ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

MINES ParisTech

École doctorale n° 432 : Science des métiers de l'ingénieur

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

l'École nationale supérieure des mines de Paris

Spécialité " Géostatistique "

présentée et soutenue publiquement par

Christian PRINS

14 Janvier 2011

Échantillonnage, simulation et estimation des gisements secondaires de diamant

Directeur de thèse : **Christian LANTUÉJOUL**
Co-encadrement de la thèse : **Wynand KLEINGELD**

Jury

M. Roussos DIMITRAKOPOULOS, Professeur, PhD, Université Mac Gill de Montréal	Rapporteur
M. Jean-Jacques ROYER, Professeur, Thèse d'état, CNRS	Rapporteur
M. Jean-Paul CHILÈS, Directeur de recherche associé, HDR, MINES ParisTech	Examinateur
M. Wynand KLEINGELD, Consultant, PhD, Indépendant	Examinateur
M. Christian LANTUÈJOUL, Professeur, HDR, MINES ParisTech	Examinateur

T H È S E

MINES ParisTech
Centre de Géosciences - Équipe Géostatistique
35 Rue St Honoré, 77300 Fontainebleau, FRANCE

Declaration

I declare that this thesis is my own work, except where otherwise acknowledged or referenced in the text. It is being submitted for the Degree of Doctor of Philosophy at the Ecole des Mines des Paris. It has not been submitted before for any other degree or examination at any other university.

C. F. Prins
January 2011

* * * * * * * * * *

Dedicated to my parents:

"Thank you for a lifetime of education."

Acknowledgements

Under the management of Dr. Wynand Kleingeld and through financial backing from the De Beers Family of Companies, an opportunity was created for a research team to look at the manner in which sampling, estimation and exploration of diamonds are done to achieve results faster, appropriately accurate and with less associated cost. A department dedicated to do research and development was set up in Wells, England, which enabled this to be done.

During this research, many hours of stimulating arguments and discussion were held with Dr. Wynand Kleingeld to whom I am indebted for many years of opportunities to grow, argue and experience different aspects of mineral resource evaluation. I am also indebted to Johan Ferreira (De Beers - retired), Christian Lantuéjoul (Ecole des Mines de Paris: Centre de Géostatistique) and Sean Duggan (previously from De Beers/now a director of Z Star Mineral Resource Consultants) for many years of association and problem solving in the field of diamond exploration and estimation.

Acknowledgements are given to De Beers Marine (Pty) Ltd, De Beers Marine: Namibia, De Beers Canada Inc (Exploration) and Namdeb Diamond Corporation (Pty) Limited for making data and images available.

I would like to express my gratitude to the De Beers Family of Companies for the opportunity, generous time and financial assistance which allowed me to complete this research.

* * * * * * * * * *

Contents

Preamble

"There are many facets to a mining operation which are amenable to change. The only facet which is unalterable is the ore deposit itself and the characteristics endowed by nature. Any change in the expectation of what the deposit can yield is related to changes in knowledge and understanding of the deposit[1]".

This knowledge and understanding is gained through measurements obtained from a sampling process of which Pierre Gy makes the following statement: "Quality estimation is a chain and sampling is its weakest link" [2]. The question is then, if sampling is the weakest link, how can one ensure that enough information of sufficient quality is obtained to enable informed decisions to be made, with the practical constraints and available technology of modern sampling campaigns?

Furthermore, during recent years it has become evident that when secondary diamond deposits are considered, historical practices of sample collection through many samples or large samples are at times not financially viable anymore. Therefore the challenge (and reason for this research), is to identify the point where the risk in decision making is at an acceptable level based on a minimum amount of sampling.

In the chapter entitled "Sample optimisation for grade when doing exploration", a method is described that circumvents requirements of well established principles of sampling practices and mathematical interpretation, but in this case decisions about sampling have to be made where insufficient information is available to establish an underlying statistical or interpretive model. It is recognised that mathematically sound models are the correct way to proceed, but the case study will discuss and demonstrate the need for such an analysis constrained by the lack of available information. It would be reckless to imply that the decision or advice will be entirely correct and unchallenged, but rather some indication of a "more correct way to proceed", is often what is required at this stage of a project, with a full understanding of the risk(s) associated and assumptions made with such advice. In these cases, sampling and estimation decisions are made under conditions where the subsequent actions are irreversible and any process that can express sampling efficiency would be of benefit.

The figure below graphically depicts the area of study for this research as that part of a project development cycle where there is an increase in the information available when migrating from speculative interpretation to quantitative measurement.

[1] *La Geostatistique Pour Des Variables Discretes* by Dr. WJ Kleingeld, Phd thesis, Centre de Géostatistique, Fontaineblaeu, 1987

[2] *Sampling for analytical Purposes* by P Gy, John Wiley & Sons, 1998, p.12

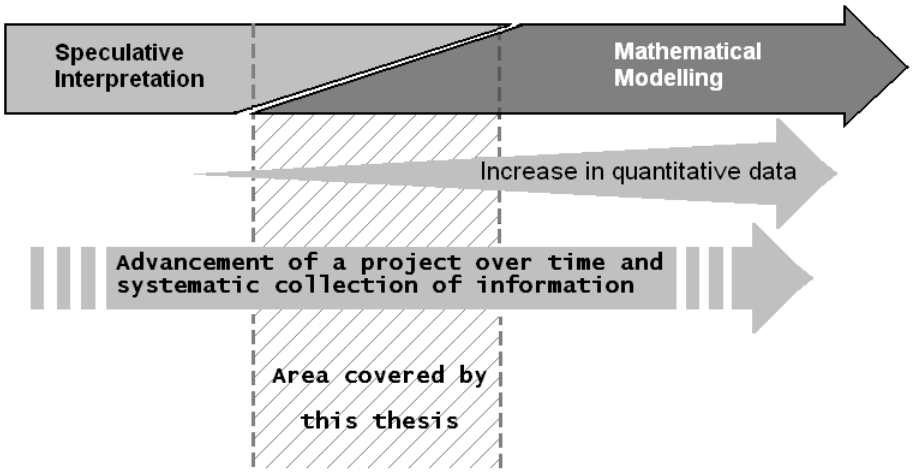


Figure 1: Migration from speculative interpretation to rigourous analysis

It is at this stage of projects where the research investigates methods to guide the collection of information on which future, more rigourous statistical models will be based.

Where sufficient information is available, the research addresses new approaches to the interpretation of exploration data and for the marine environment, to determine optimal sampling efficiency and appropriate estimation methodology. The research of each chapter is accompanied with case studies and examples.

* * * * * * * * * *

Chapter 1

Introduction

1.1 Les gisements alluvionnaires de diamants

Les cheminées de kimberlite constituent la source primaire de diamants. Apparues il y a des millions d'années par remontée de fluide magmatique, elles furent peu à peu érodées. En Afrique du Sud, le matériel érodé fut charrié le long de rivières pour atteindre la côte où il fut soumis à l'action des vagues et à un courant venu de l'Antarctique (le Benguela). Les diamants se déposèrent finalement pour constituer divers types de gisements alluvionnaires (fluviatiles, cotiers, sous-marins, éoliens). A noter que durant le transport, les pierres de mauvaise qualité furent peu à peu détruites, de sorte que la qualité moyenne des diamants qui ont subsisté est la plus élevée au monde.

De leur coté, les gisements secondaires maritimes se sont formés par phases successives de transgression et de régression, entrecoupées de périodes d'aggradation. Des tempêtes ont dispersé les pierres dans de vastes étendues, provoquant leur concentration dans certaines structures topographiques associées à la géologie. Des inondations périodiques du bassin de la rivière Orange ont déplacé le gravier diamantifère de façon discontinue vers l'océan.

La combinaison des forces d'érosion et des caractéristiques du socle (dureté, type de roche pente etc.) a entraîné la formation d'un certain nombre de structures topographiques telles que marmites de géant et ravineaux, qui sont autant de pièges pour la concentration des pierres.

Les gisements alluvionnaires ainsi formés le long de cote d'Afrique du Sud et de Namibie sont uniques en leur genre et de classe mondiale.

La partie terrestre de ces gisements est couramment exploitée (plages, rivières) avec des équipements de grande taille. En sous-marin, les gisements sont exploités entre 100m et 140m de fond par des bateaux équipés d'une usine de traitement tandis que plus près du rivage, le travail est effectué par des plongeurs jusqu'à une profondeur de 30m.

Ces gisements pouvant être de très grande taille, leur exploitation systématique n'est pas envisageable. Leur mise en valeur doit être faite de façon sélective, ce qui passe par une identification préalable de leurs parties riches par échantillonnage.

Tous ces types de gisements, auxquels on peut encore adjoindre les terrils, sont répertoriés comme gisements secondaires, simplement parce qu'ils ne constituent pas la source première de diamants. C'est sur eux que va porter notre travail de recherche.

1.2 Importance d'un échantillonnage fiable

L'estimation précise des ressources minérales nécessite un échantillonnage. Il importe de le faire soigneusement, afin de réduire les risques d'erreur et d'améliorer la prise de décision lors des phases de planification et d'exploitation. A chaque échantillonnage des objectifs clairs doivent être assignés.

Une campagne d'échantillonnage peut avoir plusieurs objectifs, parfois pas toujours compatibles. Considérons par exemple un échantillonnage d'exploration. Pour prouver l'existence de la minéralisation et se faire une idée des teneurs en place, certains géologues n'hésiteront pas à entreprendre un échantillonnage préférentiel. Ce mode d'échantillonnage n'est pas adapté à l'estimation des ressources, pour laquelle une campagne d'échantillonnage non biaisé est nécessaire, avec des prélèvements effectués sur toute l'étendue du gisement et recouvrant tous les environnements géologiques, ceci afin d'obtenir une teneur représentative.

Cela nous amène à nous demander ce qu'on appelle échantillonnage représentatif. A première vue, cette question paraît simple. Toutefois, les quelques définitions de l'échantillonnage parues dans la littérature ne sont pas identiques. Le mot change de sens selon qu'il émane d'une référence statistique, géologique ou bien d'exploitation minière. En voici quelques exemples[1]. L'échantillonnage est

- l'art de prélever une petite partie d'une entité inconnue par des moyens divers, dans le but de montrer en quoi consiste le tout, à un niveau de confiance près;
- la sélection d'une partie statistique significative, analysée par une méthode industrielle reconnue ou bien mesurée selon une technique appropriée pour estimer les caractéristiques d'un groupe ou d'une population plus grande;
- la création d'un ensemble d'observations enregistrées dont une partie constitue un échantillon statistique. Un ensemble d'observations ou d'échantillons recueillis lors d'une campagne géologique (définitions tirées du glossaire géostatistique d'Olea);
- l'opération qui consiste à prélever une partie représentative et apte à l'analyse d'un tout bien plus important, de telle façon que la proportion et loi de la caractéristique à mesurer soient, dans des limites raisonnables, les mêmes dans la partie prélevée (l'échantillon) et dans le tout (la population)[2]

Pour échantillonner les gisements de diamants et évaluer leurs ressources, c'est cette dernière définition qui est la plus complète et qui sera ultérieurement discutée. Elle est logique et son mode d'utilisation est intuitif. L'étude précise de cette définition fait bien ressortir la complexité d'un échantillonnage. Sa mise en oeuvre ne doit pas être sous-estimée.

1.2.1 Les différentes composantes de l'échantillonnage et l'introduction de l'erreur d'échantillonnage

La définition de l'échantillonnage selon Covacic introduit un certain nombre d'éléments-clés:

[1] *Guidelines for Mineral Resource Management* by SP Duggan, Internal De Beers Report, April 2008
[2] *Sampling Techniques*, Mining Magazine by D Covacic & J Clarke, October 1990, p270.

L'échantillonnage est défini comme l'opération$^{(A)}$ *qui consiste à prélever*$^{(B)}$ *une partie représentative*$^{(C)}$ *et apte à l'analyse*$^{(D)}$ *d'un tout bien plus important*$^{(E)}$*, de telle façon que la proportion et la loi*$^{(F)}$ *de la caractéristique à mesurer*$^{(G)}$ *soient, dans des limites raisonables, les mêmes*$^{(H)}$ *dans la partie prélevée et dans le tout*$^{(I)}$*.*

Cette définition va être critiquée terme à terme pour en faire ressortir les difficultés et amener progressivement l'idée d'erreur d'échantillonnage.

(A) **défini comme l'opération:** les échantillons des dépôts secondaires sont nécessairement de grande taille et par conséquent traités dans des unités d'exploitation. Les teneurs obtenues ne sont pas toujours parfaitement restituées ce qui soulève le problème de l'intégrité de l'échantillonnage;

(B) **prélever:** lorsque l'on effectue des sondages ou bien lorsque que l'on lève des tranchées dans des matériaux peu consolidés, des effondrememements peuvent toujours survenir. Ils entrainent une contamination des échantillons, rendant difficile l'évaluation exacte de leur volume et de leur poids;

(C) **partie représentative:** que signifie-t'elle, et comment la mesure-t'on? Il est bien connu que des échantillons de petite taille ont des teneurs très fluctuantes. Ce phénomène est encore accentué par le caractère discret de la minéralisation: même prélevés dans des zones minéralisées, de tels échantillons peuvent ne renfermer aucune pierre. C'est la raison pour laquelle on préfère recourir à des échantillons de grande taille, au risque de perdre de l'information au niveau de la microstructure;

(D) **et apte à l'analyse:** usuellement de quelques tonnes, le poids des échantillons peut s'élever à 40000 tonnes. Dans le cas des terrils, il peut même exceptionnellement monter à 150000 tonnes. L'impossibilité pratique de traiter des échantillons de cette taille crée un problème de représentativité;

(E) **d'un tout bien plus important:** ceci suppose les dimensions du "tout" connues. En réalité, le domaine géologique est difficile à délimiter. Il arrive parfois qu'il soit un assemblage de sous-domaines géologiques;

(F) **de telle façon que la proportion et la loi:** ceci sous-entend que des modèles statistiques peuvent être utilisés pour résumer l'information des données. Il faut souligner qu'un échantillon n'est qu'une version de la réalité prélevée dans un certain environnement géologique. Il faut donc s'attendre à des fluctuations statistiques;

(G) **la caractéristique à mesurer:** des erreurs de mesure apparaissent suite à la perte de certains diamants lors du traitement minéralurgique de l'échantillon;

(H) **les mêmes dans des limites raisonnables:** le test d'identité et le choix des limites dépendent tous deux du niveau de risque jugé acceptable. Les résultats obtenus ne sont jamais parfaitement corrects et restent toujours sujets à interprétation.

(I) **dans la partie prélevée et dans le tout:** les résultats peuvent être biaisés si les échantillons ne sont pas prélevés dans tout le gisement.

C'est en prenant conscience de ces différentes composantes que l'on pourra s'assurer de la bonne qualité et de la fiabilité des échantillons que l'on sera amené à utiliser dans des programmes d'estimation ou de simulation.

Dès qu'un échantillon est prélevé, il est intégré au jeu de données et devient utilisable pour tout le reste du projet. La Figure 1.1 montre l'introduction du jeu de données dans le cycle d'estimation des ressources minérales. Cette figure montre que l'échantillonnage initial a l'impact le plus important. Pour cette raison, il est essentiel qu'il soit effectué correctement. La tâche est d'autant plus délicate que c'est précisément au tout début du projet qu'il est le plus difficile de bâtir un plan d'échantillonnage.

La discussion précédente et l'introduction potentielle de l'erreur d'échantillonnage ne doit pas laisser l'impression qu'il est impossible de bien échantillonner. Il s'agissait plutôt d'une incitation à échantillonner soigneusement de façon à garantir la correction des échantillons.

1.3 Problèmes d'échantillonnage

Pour faire face au besoin permanent d'estimer les teneurs, il est important de chercher à optimiser l'échantillonnage. Les échantillons ainsi obtenus sont plus fiables et plus représentatifs ce qui fait gagner en rentablité au moment de l'exploitation.

Par exemple, il a été constaté que les gisements sous-marins de grande profondeur (de 100 à 140m sous le niveau de la mer) se découpent en grandes zones géologiquement homogènes. Des travaux de recherche furent entrepris entre 1990 et 1996 pour optimiser la taille des prélèvements et leur espacement. Les recommandations tirées de cette recherche sont toujours en vigueur, au bénéfice de la société depuis 15 ans (voir Annexe A).

De nos jours, l'exploration a lieu dans des zones de teneurs de plus en plus faibles, pour lesquelles l'estimation des teneurs soulève de grandes difficultés d'échantillonnage. Dans certains cas, l'information disponible est très faible, sinon nulle. De nouvelles méthodes doivent être mises au point pour optimiser l'effort d'échantillonnage.

Dans ces zones de contours géologiques peu établis, soit en raison d'une information géologique insuffisante, ou bien parce que les teneurs disponibles n'apportent rien à leur détermination, des hypothèse statistiques telles que la stationnarité ne peuvent être postulées. Dans un tel cas, tout résultat, même approximatif, peut servir de guide à une campagne d'échantillonnage.

La mise en place d'une procédure d'échantillonnage soulève deux types de questions, celle de sa paramétrisation (taille et espacement) et celle de sa correction (comment prélever des échantillons de bonne qualité?). Les quelques exemples qui suivent montrent que les réponses à apporter sont parfois subtiles et difficiles.

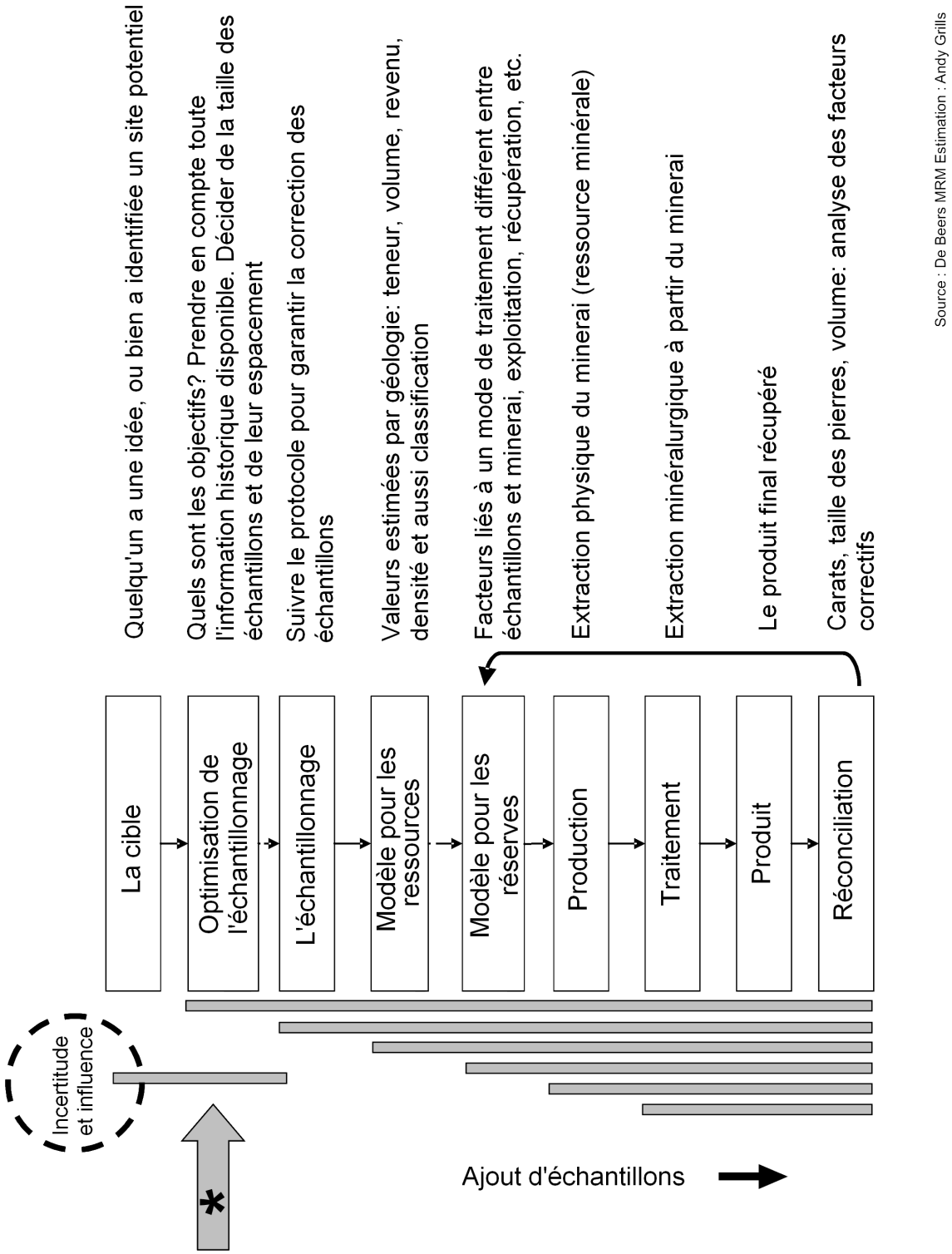


Figure 1.1: Introduction des données d'échantillonnage dans le cycle d'évaluation des ressources minérales

1.4 Exemples de problèmes d'échantillonnage

Dans ce travail de recherche, on fera fréquement référence aux foreuses vibrantes et rotatoires utilisées pour prélever les échantillons de grand fond (cf. Figure 1.2).

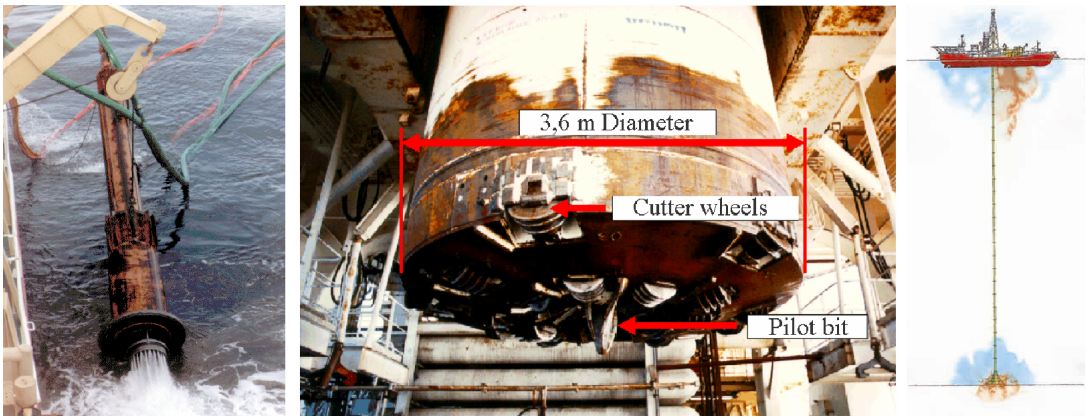


Figure 1.2: Outils d'échantillonnage en sous-marin: foreuse vibrante à gauche et foreuse rotatoire à droite

– Foreuse vibrante (cf. Figure 1.2 à gauche): cet appareil prélève des échantillons de 0.72m^2 en faisant rebondir la tête de forage (dans sa chemise) sur le fond. L'opération se poursuit jusqu'à ce que le socle soit atteint ou bien que la remontée du gravier à bord du bateau s'interrompe, ce qui peut être vu sur les écrans de contrôle;

Foreuse rotatoire (cf. Figure 1.2 à droite): Il s'agit d'un appareil de marque Wirth équipé de molettes à disques et d'une tête de guidage. il est rattaché au navire par une série de tuyaux qui fournissent la torsion nécessaire à la rotation de la tête de forage, et assurent la remontée du gravier.

Les matériaux remontés du fond sont traités à bord pour en extraire les diamants, puis rejetés en mer.

Prélèvement d'échantillons de grande taille pour récupérer suffisamment de pierres dans les zones de faible teneur

Le problème posé était d'échantillonner le gisement paléofluvial de Sendelingsdrift en Namibie. Des échantillons allant jusqu'à 40000 tonnes ont été prélevés pour recueillir un nombre suffisant de pierres en vue de l'estimation des teneurs. Le nombre de pierres ainsi collectées a varié de 121 (pour un échantillon de 37000 tonnes) à 2944 (pour un échantillon de 21000 tonnes).

Au moment d'envisager de tels échantillons, la question s'est posée de savoir s'il valait mieux prélever de nombreux petits échantillons plutôt que quelques gros. La conclusion qui prévalut, quelques gros échantillons, fut obtenue en recourant à une version simplifiée des méthodes de simulation alors en usage.

Figure 1.3: Taille et emplacement des échantillons du gisement de Sendelingsdrift sur la rivière Orange, à la frontière entre l'Afrique du Sud et la Namibie

Huit échantillons de $40m \times 40m$ ont été creusés pour représenter le gisement, d'environ $8km \times 1km$. Le nombre d'échantillons a été soigneusement calculé et leur emplacement a été déterminé à partir de la géologie du terrain et de son interprétation.

Formation d'échantillons "vermiculaires" par forage rotatoire

L'efficacité de la foreuse rotatoire a été testée avec ou sans guidage. Sans ce système, la tête de forage ne pénètre pas correctement dans le terrain comme le montre la figure ci-dessous:

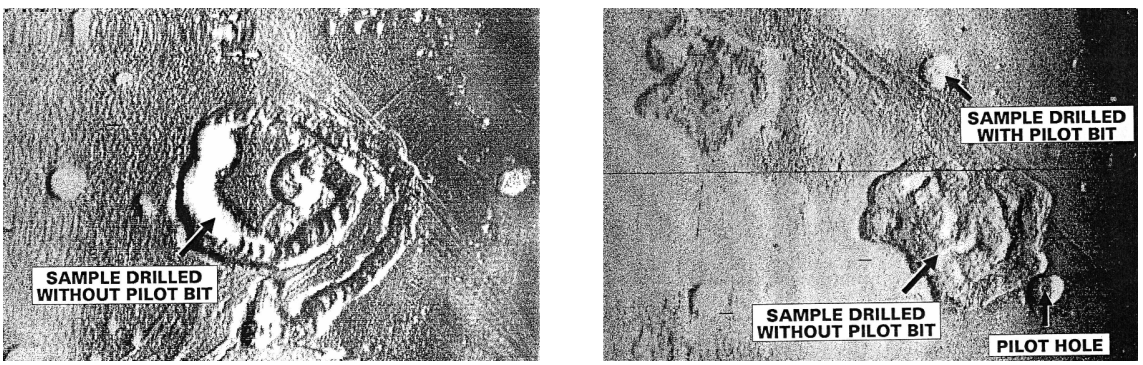


Figure 1.4: Ces deux exemples montrent la prise d'échantillons par forage rotatoire avec ou sans tête de guidage

Ainsi, les échantillons obtenus sans guidage n'auraient pas été représentatifs. La tête de guidage a dû être dimensionnée pour lui permettre de fonctionner quelle que soit la dureté du terrain. Une grande taille marcherait très bien sur des terrains meubles mais ne conviendrait pas à des terrains plus durs. Une taille optimale fut déterminée par essai et erreur.

Surpénétration des échantillons

Au large des côtes sud africaines, les échantillons ont été prélevés par foreuses vibrantes et rotatoires pour l'estimation des teneurs. Il se trouve que l'échantillonnage perd en efficacité lorsque le terrain passe rapidement de l'argile (dure) au sable (plus mou). Dans un terrain meuble, on observe une contamination de l'échantillon par la succion de matériaux du voisinage. Pour réduire cette nuisance, plusieurs configurations de chemisage ont été testées sur les foreuses. La figure 1.5 illustre la perte de représentativité observée quand on passe d'un terrain argileux à un terrain sablonneux.

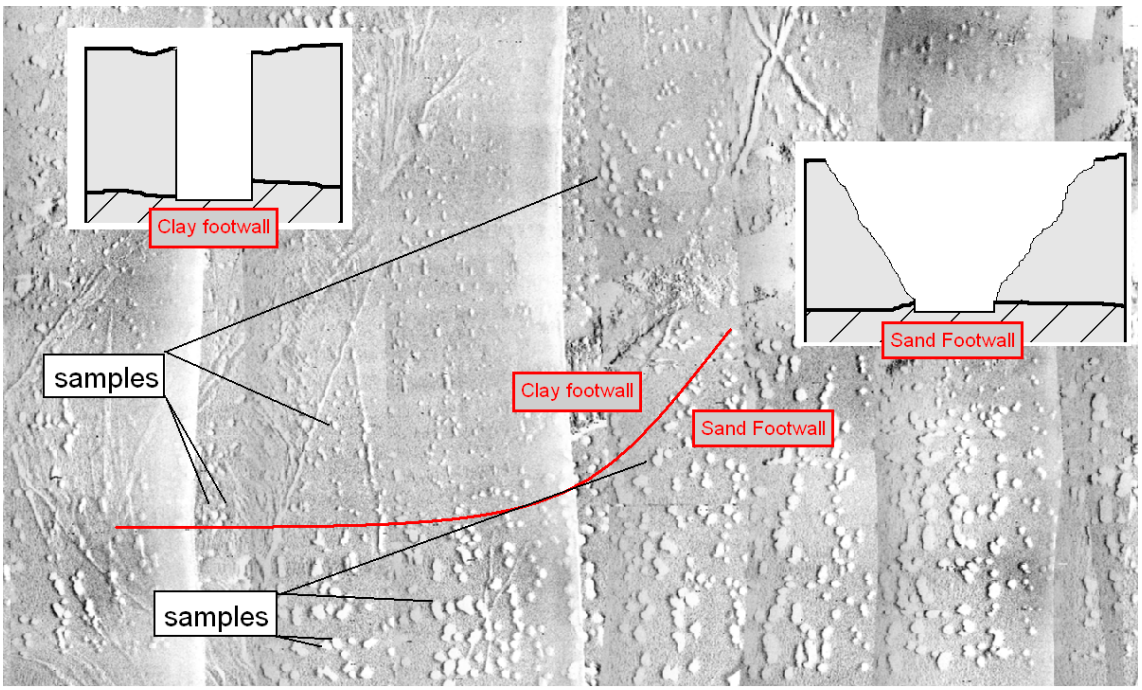


Figure 1.5: Effet du terrain sur la correction des échantillons: passage d'un terrain argileux à un terrain sablonneux

Dans certaines parties du gisement, la correction des échantillons est exemplaire, telle que celle montrée en Figure 1.6:

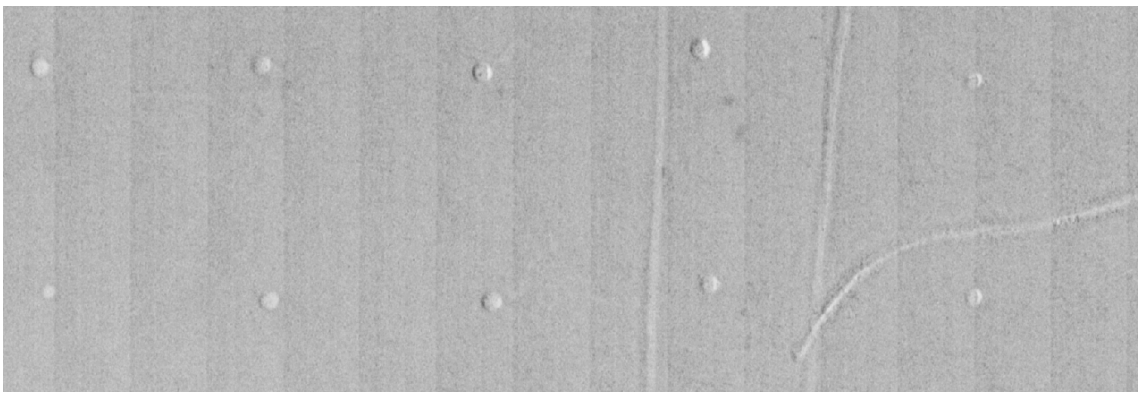


Figure 1.6: Echantillons de $10m^2$ d'excellente qualité (cela ne montre toutefois pas si les échantillons ont bien pénétré dans le terrain)

Le degré de contamination n'est malheureusement pas toujours mesuré sur tous les échantillons. Ses conséquences sur les résultats d'estimation sont difficilement évaluables.

Prise en compte du fléchissement de la tête de guidage

On a vu que des têtes de guidage de longueurs différentes ont été testées au moment des premières utilisations des foreuses rotatoires. Un fléchissement de ce système a été observé pendant la prise des échantillons, conduisant à des décalages de $40cm$. Compte tenu de la rotation de la tête de forage, l'échantillon obtenu est de plus grande taille. Alors que $40cm$ ne semble pas bien important par rapport au diamètre du forage ($3.6m$), le changement de surface n'en est pas moins conséquent. La surface de l'échantillon passe de $10.18m^2$ à $15.20m^2$, soit une différence de 49%. Ce problème fut très vite identifié et rectifié.

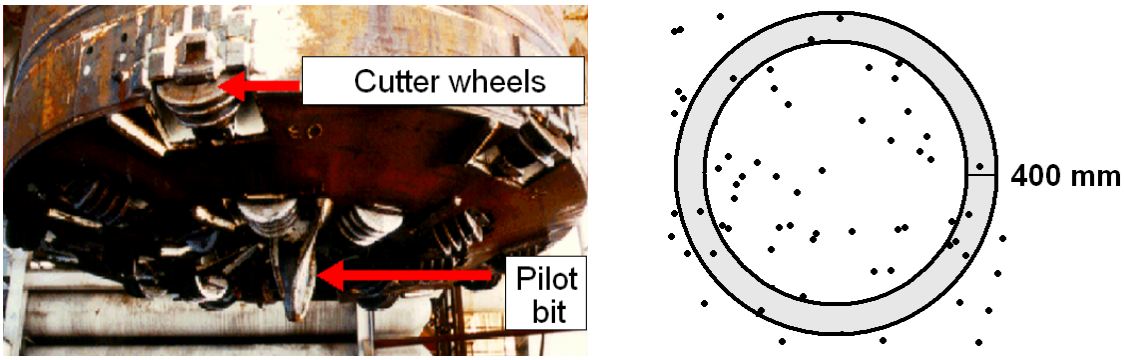


Figure 1.7: Effet du fléchissement de la tête de guidage

Même sans fléchissement de la tête de guidage, la rotation de l'outil de forage crée un biais similaire sur les résultats d'échantillons. Un problème plus classique est la remontée des matériaux des abords du trou de forage. Elle survient en terrain meuble. La difficulté qu'il y a de nettoyer le fond de l'échantillon est aussi l'une des causes de l'inefficacité de l'échantillonnage. Ces problèmes sont continûment réexaminés à chaque fois que sont développés de nouveaux systèmes de remontée des matériaux ou bien de nouveaux chemisage pour les foreuses, pour améliorer l'efficacité aussi bien de l'échantillonnage que de l'exploitation. Une tête de guidage raccourcie, comme celle de la figure 1.7 ci-dessus, est d'utilisation courante.

Effet de "circulation" des pierres

Durant le processus d'échantillonnage, un certain nombre de fausses pierres, de même densité et de même luminescence que de vrais diamants, sont insérées dans les matériaux remontés. La proportion de fausses pierres récupérées fournit un estimateur de l'efficacité de l'usine de traitement. Un tel exercice a fait apparaitre une circulation importante des pierres. Une correction des données a été tentée pour pouvoir les utiliser dans un programme de simulation. Pour cela, on a enregistré la date d'obtention des échantillons, ce qui a permis de les classer par ordre chronologique. Les pierres ont été réattribuées à leur échantillon d'origine en tenant compte des performances de l'usine de traitement. Pour cela, désignons respectivement par X_n et O_n les nombres vrais (mais inconnus) et observés (donc connus) de pierres de l'échantillon

n. En s'appuyant sur la récupération des fausses pierres, on fait l'hypothèse selon laquelle les pierres de O_n viennent à 85% de l'échantillon n, à 12% de l'échantillon $n-1$ et à 3% de l'échantillon $n-2$:

$$O_n = 0.85X_n + 0.12X_{n-1} + 0.03X_{n-2}$$

Le nombre exact de pierres de l'échantillon n s'écrit alors:

$$X_n = \frac{O_n - 0.12X_{n-1} - 0.03X_{n-2}}{0.85} \tag{1.1}$$

On a de même

$$X_{n-1} = \frac{O_{n-1} - 0.12X_{n-2} - 0.03X_{n-3}}{0.85} \tag{1.2}$$

En injectant alors la formule (1.2) dans l'équation (1.1), puis en éliminant les termes d'amplitude négligeable, on parvient à écrire X_n sous la forme

$$X_n \approx 1.18O_n - 0.17O_{n-1}$$

Ce modèle, qui vise à restituer leur valeur vraie aux échantillons, n'est qu'un effet indésirable du manque d'efficacité de l'échantillonnage.

Ecarts entre statistiques d'échantillons et de production

Au cours d'un exercice de simulation entrepris par Prins [47,48] dans le but de tester la possibilité de recourir à des prélèvements de plus grande taille pour échantillonner la zone minière Atlantic 1 (au large des cotes de Namibie), il s'est trouvé que les deux sources d'information disponible, à savoir les données de forage vibrant et les données de production, fournissaient des statistiques notablement différentes.

Les données de production étaient considérées comme plus fiables, mais les paramètres du modèle à simuler devaient être spécifiés à un support équivalent à celui des forages. Une déconvolution du support de production (blocs de $50m \times 50m = 2500m^2$) au support des forages n'aboutit pas.

De leur coté, les données fournies par forage vibrant, seule autre source d'information, manquaient de fiabilité pour déterminer les paramètres du modèle à simuler. En fin de compte, un processus itératif par essai et erreur a été mis en place.

Des valeurs de paramètres du modèle sont proposées en entrée. Elles ne sont acceptées que si elles fournissent des statistiques de blocs simulés comparables à celles des blocs de production. Quelques itérations suffisent pour ajuster les paramètres (voir schéma de la Figure 1.8).

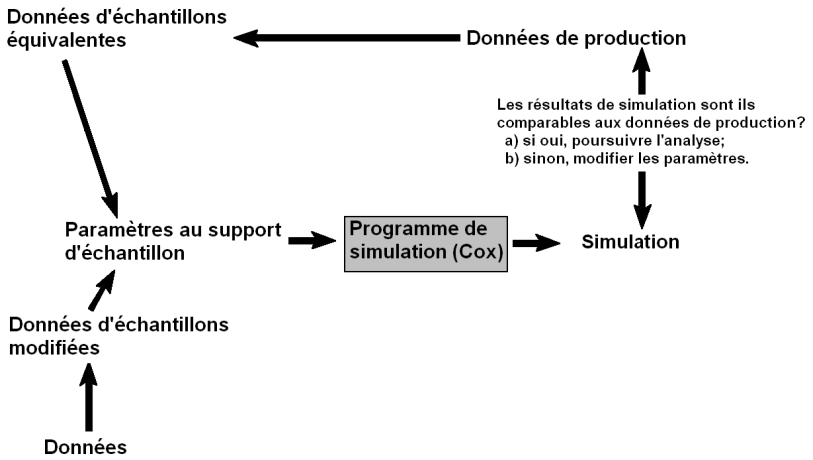


Figure 1.8: Des valeurs sont attribuées itérativement aux paramètres du modèle de simulation jusqu'à ce que les statistiques des blocs simulés soient comparables à celles des blocs de production

Par cet algorithme, les paramètre du modèle à simuler purent être déterminés avec succès.

Complexité des environnements géologiques sous-marins

Certains sites sous-marins peuvent être de géologie compliquée. La figure ci-dessous montre la complexité d'une interprétation géologique régionale. Chaque couleur représente un faciès géologique différent.

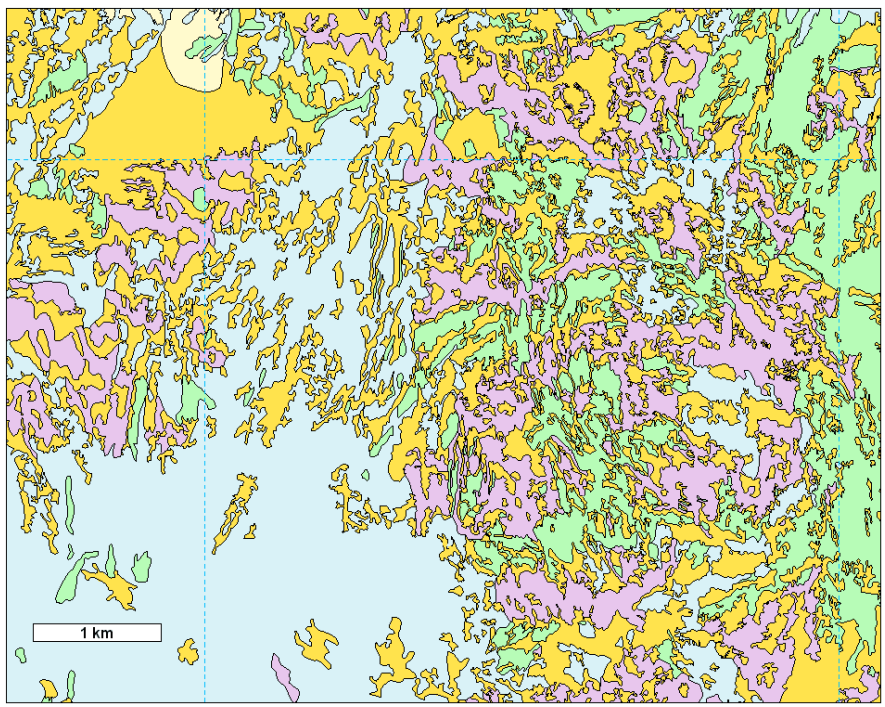


Figure 1.9: Système complexe de faciès géologiques et de pièges

Cette zone n'a rien d'homogène et son échantillonnage constitue une véritable gageure. Le caractère directionnel de la géologie suggère d'effectuer un krigeage à voisinages orientés (voir Annexe D).

Prise en compte d'echantillons de différents supports

Les données sont souvent constituées d'échantillons de tailles et de configurations différentes. Ceci soulève le problème de leur prise en compte conjointe dans des procédures d'estimation ou de simulation. L'exemple de la Figure 1.10 montre des trous et des tranchées creusés près d'Oranjemund (Namibie). Ces différents mode d'échantillonnage sont apparus au cours du temps, au fil des évolutions technologiques et des fluctuations économiques.

Figure 1.10: Exemples de prélèvements par trous et par tranchées

Le même type de problème survient lorsque l'on cherche à optimiser une campagne d'échantillonnage en milieu marin à partir d'échantillons de tailles différentes ($0.72m^2, 10m^2$ et $2500m^2$). En raison du caractère discret des données, leur prise en compte à des supports différents nécessite l'introduction de méthodes spécifiques.

Les horizons cimentés et leurs conséquences sur l'estimation des teneurs

Dans certains gisements marins, l'environnement de dépôt a cimenté le matériel, piégeant les pierres dans des matériaux consolidés. La teneur des zones cimentées est la même qu'aux alentours, mais il n'est pas toujours possible de les échantillonner et il est difficile d'en récupérer les pierres. La partie terrestre de la mine d'Elisabeth Bay, détenue par Nambed, comporte d'importantes zones cimentées. Son échantillonnage est montré en Figure 1.11.

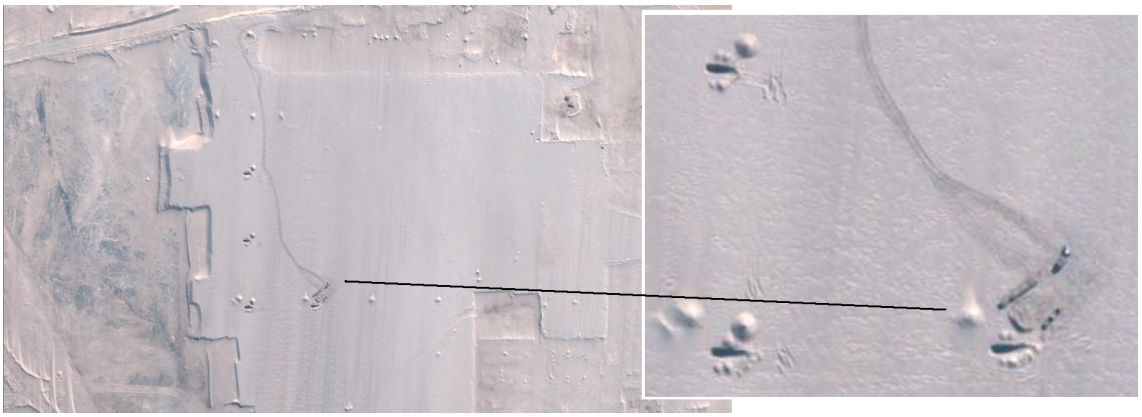


Figure 1.11: Echantillonnage de la mine d'Elisabeth Bay (Namibie)

Les outils d'échantillonnage standard ne parviennent pas toujours à pénétrer dans le matériel cimenté, de sorte que les épaisseurs mesurées sont peu fiables. Les dimensions des zones cimentées étant assez variables, il a fallu, en de nombreux endroits, formuler des hypothèses sur leur épaisseur sans tenir compte des horizons stratigraphiques. Cet ensemble de circonstances, qui affecte les mesures de teneur et d'épaisseur, a des répercussions importantes sur la précision des estimateurs.

Mesures de teneur dans des environnements érosifs très irréguliers

Dans les gisements de bord de mer, les ravineaux et les marmites de géant sont susceptibles de contenir un grand nombre de pierres, mais leur emplacement erratique fait qu'il est difficile d'en prédire la position. Dans certains cas, l'appareillage d'échantillonnage est suffisamment agressif pour forer dans le socle. Si ce n'est pas le cas, la représentativité des échantillons doit être remise en question, tout particulièrement si l'échantillonnage révèle une teneur plus faible que prévue. Des exemples de pièges bien dégagés sont montrés en Figure 1.12.

Figure 1.12: Ravineaux et marmites de géant et leur implantation erratique

Pour des projets d'ampleur plus marginale, il convient de prélever un certain nombre d'échantillons pour déterminer le degré de minéralisation des ravineaux. Des travaux sont poursuivis, aussi bien en interne[3] qu'à l'extérieur de la de Beers[4], pour estimer la pénétration des ravineaux en fonction de la taille des l'échantillons, l'objectif étant d'attribuer une précision aux estimations de teneurs dans les zones érodées.

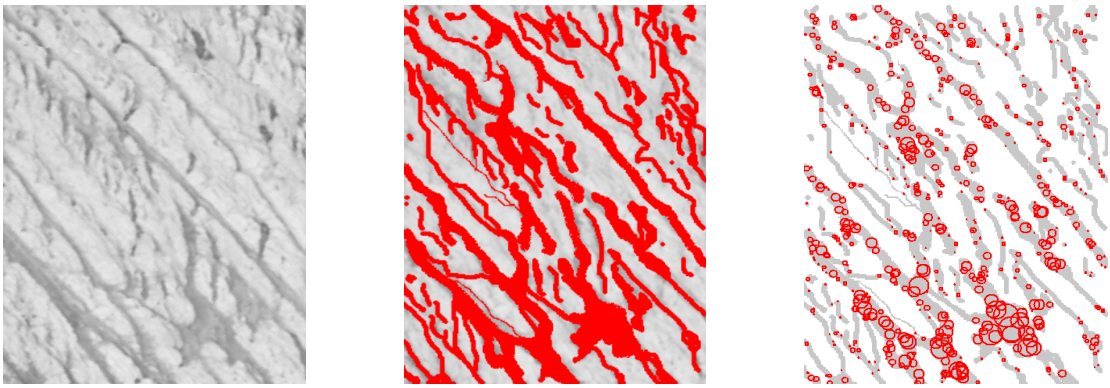


Figure 1.13: Etude de l'aptitude qu'ont différents échantillons à pénétrer dans des ravineaux en fonction de leur taille. De gauche à droite: l'image originale, l'image interprétée et quelques positions et tailles de pénétration

Pour quelques projets d'intérêt économique marginal, il est important de savoir comment varient les teneurs estimées en fonction du taux de pénétration dans les ravineaux.

En résumé:

L'échantillonnage et l'exploitation de gisements secondaires requièrent tout un savoir faire pour parvenir à sans cesse optimiser les paramètres d'échantillonnage, de simulation, d'estimation et d'exploitation. Les exemples montrés ci-dessus donnent une idée de la diversité des problèmes et des environnements pour lesquels des travaux d'échantillonnage et d'estimation ont été entrepris. Au travers de cette amélioration sans relâche des procédures d'échantillonnage et d'estimation, différents thèmes de recherche ont été identifiés.

1.5 Recherche entreprise

L'échantillonnage et la simulation des gisements secondaires sont à la base de deux chapitres de cette thèse. Deux autres chapitres sont consacrés à l'echantillonnage des terrils ainsi qu'à l'analyse de données indicatrices de la présence de kimberlite dans le but d'identifier les zones à explorer en priorité. Ces chapitres ne sont pas directement liés entre eux. Ils ont été ordonnés de façon à couvrir tout d'abord les problèmes d'exploration, puis d'optimisation de l'échantillonnage des gisements secondaires, et enfin d'échantillonnage des terrils.

[3] *Determining largest circular sample size across a network covering a gullied area* par C.F. Prins en mars 2009

[4] *Investigating gully penetration using different sample sizes*, thèse de master défendue par J. Jacobs en septembre 2009 à l'université du Witswaterrand

1.5.1 Regroupement des kimberlites et identification des cibles d'exploration au moyen d'indicateurs

L'exploration des kimberlites est l'une des grandes priorités de la de Beers. Toute méthode fournissant des résultats à moindre coût ou plus efficacement sera profitable à la compagnie. Des jeux de données ont été confiés aux départements d'exploration de la de Beers dans le but de les aider à trouver des méthodes (géo)statistiques qui pourraient servir à trouver des kimberlites. Il existe des emplacements de kimberlites connus en association avec les données des cratons sud-africains et canadiens. Ils sont utilisés pour construire des modèles et aussi dans des cas d'étude.

Travaux de recherche:

Cette recherche comprend plusieurs parties:

Premièrement, une cartographie des emplacements des kimberlites dans le craton sud-africain a fait apparaître la possibilité d'un regroupement en amas. L'existence de tels amas a déjà été étudiée en interne par Barnett [3], mais les méthodes préconisées et implémentées dans cette thèse sont résolument nouvelles. La première partie de cette recherche a pour but de confirmer cette impression visuelle d'amas, puis à déterminer leur extension moyenne. La connaissance des tailles d'amas sert à:

a) demander une licence d'exploitation. Celle-ci donne à la compagnie la possibilité d'explorer un territoire pendant un certain temps (dépendant du pays), puis de n'en garder que des parties choisies selon le principe "on exploite ou bien on laisse". La connaissance du comportement des anomalies kimberlitiques et de la taille des amas fournit une information utile à la décision. b) délimiter les zones de prospection. Dès qu'une anomalie a été identifiée et jugée potentiellement intéressante, il est avantageux d'explorer son voisinage par voie aérienne pour maximiser les chances de trouver l'emplacement d'autres cheminées kimberlitiques.

Deuxièmement, de larges étendues de l'Afrique du Sud et du Canada ont fait l'objet de campagnes de mesures d'indicateurs minéraux. Ces indicateurs sont bien plus fréquents que les diamants et plus facilement décelables au voisinage et à l'intérieur d'une cheminée kimberlitique. Le profil d'un indicateur minéral varie selon le type de kimberlite et des caractéristiques de ses particules (taille, dureté et densité) qui les rendent aptes à migrer sur de plus ou moins longues distances. Leur observation en grande proportion indique la proximité d'une kimberlite. L'analyse de ces données est une gageure car de nombreux facteurs affectent ces indicateurs minéraux. C'est en combinant les règles de migration, les connaissances géologiques et les données que l'on se donne les meilleures chances d'identifier les anomalies kimberlitiques. Le domaine sur lequel a porté l'analyse comprend plusieurs régions de géologies différentes, et les minéraux indicateurs ont été sujets à diverses conditions climatiques et à de nombreuses forces d'érosion. Leur proportion varie en fonction du type de kimberlite d'où ils proviennent. Des techniques statistiques de data mining ont été mises en jeu pour élaborer une méthode d'analyse permettant d'identifier les zones de grand potentiel.

1.5.2 Simulation à la main de gisements complexes de type placers

Etant exploités de façon sélective, les placers doivent sans cesse être échantillonnés. En présence de données abondantes, des techniques de simulation spatiales telles que celles décrites en Annexe G peuvent être utilisées pour optimiser des campagnes d'échantillonnage. Il existe des situations pour lesquelles l'échantillonnage doit être optimisé sans données. C'est le cas, par exemple, d'une zone qui vient d'être identifiée et n'est pas encore échantillonnée. Sur de telles zones, on peut disposer de quelques échantillons pour en délimiter les contours et en faciliter l'interprétation géologique. A partir de cette information, un programme d'échantillonnage doit être planifié. Dans une perspective d'optimisation, il est avantageux de s'assurer que des échantillons de taille correcte et d'espacement adéquat ont été prélevés. La méthode que nous envisageons prend en compte toute l'information disponible ainsi que les interprétations des géologues. Elle permet de quantifier l'impact de différentes stratégies d'échantillonnage. Il faut souligner que le modèle ne vaut que ce que valent les interprétations des géologues, et que l'analyse ne fournit qu'une réponse relative, précise à ordre de grandeur près. En d'autres termes, cette réponse est étroitement liée à l'expérience et à la qualité des interprétations géologiques. Une fois que les premiers résultats de teneur échantillonnée deviennent disponibles, l'interprétation géologique et le processus d'optimisation peuvent être reconsidérés en recourant à des simulations spatiales. Comment faciliter cette prise de décision préliminaire sur l'échantillonnage quand les seules données disponibles sont de nature géologique? Une façon d'appréhender le modèle de minéralisation est de recourir à une esquisse du gisement au crayon. Cette esquisse est effectuée par le géologue et représente la minéralisation en teintes plus ou moins sombres selon la richesse présumée des teneurs. La qualité et la précision de l'esquisse va déterminer la validité de la simulation, et quand bien même cette méthode ne soit pas scientifiquement fondée, il n'existe aucune autre méthode pour orienter la stratégie d'échantillonnage en tout début d'exploration. Considérons l'exemple de la Figure 1.14: la présence de gravier a été détectée dans le gisement par des sondage de type Auger. Le géologue est alors capable de bâtir un modèle qui comprend l'épaisseur et la qualité du gravier, en y incorporant ses idées sur sa richesse potentielle en diamants. Le problème devient alors de savoir comment échantillonner un tel gisement esquissé: par tranchées ou bien par sondages selon une grille régulière au moyen d'une benne preneuse. La Figure 1.14 montre aussi une esquisse du potentiel en diamants réalisée par des géologues, les parties sombres correspondant aux plus fortes concentrations. L'expérience et la compétence du géologue, l'art de l'interprétation géologique et les observations visuelles ont été utilisées pour obtenir cette esquisse. Elle intègre des informations qualitatives qui échappent à toute quantification. Lorsque l'information disponible est seulement qualitative, on propose une méthode qui, partant de l'interprétation du géologue, transforme son esquisse en simulations par des moyens statistiques. Quoique rudimentaires, elles peuvent toutefois servir à mettre sur pied un premier programme d'échantillonnage.

1.5.3 Optimisation de l'échantillonnage dans les placers sous-marins

Au milieu des années 90, Kleingeld, Prins et Duggan [20,47,48] ont entrepris un travail de recherche qui leur a permis d'établir que $10m^2$ était la taille optimale de prélèvement pour estimer le nombre de pierres dans des blocs de $50m \times 50m$ du gisement Atlantic 1 avec une précision acceptable. Ce résultat fut obtenu en recourant à des simulations non conditionnelles du processus de Cox et en s'appuyant sur des données d'échantillons (de $0.72m^2$, regroupés

Figure 1.14: A gauche, la zone à échantillonner. A droite, l'interprétation géologique indiquant les zones de plus grand potentiel diamantifère. Les lignes tiretées délimitent la zone d'intérêt

par paquets de 3), et de résultats de production. Les paramètres du modèle ont été calés à l'aide de ces données numériques. La taille optimale ainsi établie, l'appareillage fut construit. Il est toujours en usage de nos jours. Depuis, les données de $10m^2$ ont prouvé leur efficacité pour l'estimation de blocs. Toutefois, dans le but de réduire les coûts d'échantillonnage et aussi d'aborder des fonds marins de géologies différentes, il a été envisagé de réexaminer les résultats de l'étude initiale et de tester la possibilité de revenir à un instrument de forage de plus petite taille.

Optimisation de l'échantillonnage par simulation conditionnelle du processus de Cox:

On sait que les données fournies par la foreuse rotatoire ($10m^2$) sont plus fiables que celles fournies par la foreuse vibrante ($2.16m^2$). La technologie progressant, il a été avancé qu'il est maintenant possible de concevoir des échantillons de plus petite taille mais fiables, qui permettraient des estimations comparables à celles fournies par les échantillons de grande taille. Ce travail de recherche a deux objectifs:

(a) utiliser les données de $10m^2$ pour corroborer les résultats obtenus durant les années 90;

(b) tester la possibilité d'utiliser des échantillons de plus petite taille, ce qui permettrait de forer à partir d'un vaisseau plus petit, incapable de supporter le poids d'une foreuse rotatoire. Dans cette recherche, on teste divers scénarios d'échantillonnage sur des simulations conditionelles du processus de Cox.

1.5.4 Echantillonnage et estimation des ressources dans les terrils

Les ressources minérales des terrils sont constituées de minerai qui a déjà été traité en usine pour en récupérer les diamants. Ceux-ci étant de taille variable, le processus d'extraction comporte

plusieurs phases de broyage permettant de réduire peu à peu la taille des morceaux de minerai. Quoique très dur, le diamant est aussi fragile et peut se briser lors d'un broyage, si ce dernier n'est pas effectué soigneusement. Se retrouvent dans les terrils les diamants non récupérés, et cela pour diverses raisons: pertes survenues durant le traitement minéralurgique, pierres non libérées ou trop petites, minerai non traité. Le terril CTP contient 120 millions de tonnes de minerai, conséquence d'une centaine d'années d'exploitation minière à Kimberley. Il importe de concevoir des techniques d'échantillonnage et d'estimation pour en évaluer la richesse. En ce qui concerne l'échantillonnage, le problème est de déterminer une taille d'échantillon représentatif ainsi que la grille d'espacements entre échantillons qu'il convient de prendre. L'objectif de cet échantillonnage est de procéder à une estimation locale par blocs en vue de la planification minière. Ces techniques doivent fournir des résultats précis car la valeur moyenne de ces ressources est certainement proche du seuil de rentabilité.

1.6 Organisation de la thèse et résumé des méthodes utilisées

Cette thèse est organisée de la façon suivante: le premier problème étudié est l'exploration de nouveaux gisements. Une technique est ensuite proposée pour simuler la répartition spatiale des pierres d'un gisement en présence de peu de données, voire en l'absence de toute donnée quantitative. L'optimisation d'une campagne d'échantillonnage est encore un autre problème, traité dans le cas spécifique d'un gisement sous-marin. Pour finir, le problème de l'échantillonnage et de l'estimation des terrils est abordé. On donne ci-dessous un résumé du contenu des différents chapitres.

Chapitres 1 et 2: introduction

Il s'agit de chapitres introductifs décrivant les points de départ de ces travaux de recherche et fournissant toute l'information de base concernant les gisements secondaires.

Chapitre 3: détection de zones d'intérêt et analyse des résultats d'exploration des kimberlites

Tout d'abord, les kimberlites aparaissent regroupées en amas. Ce phénomène est confirmé par utilisation de la fonction de Ripley. On évalue ensuite la taille des amas et la distance qui les sépare en recourant à un certain nombre d'outils métriques (distance au plus proche voisin), variographiques (variogramme d'indicatrice) et combinatoires (graphe de voisinage, graphe de Gabriel). Ces méthodes sont appliquées à plusieurs cas d'étude et les résultats commentés. Vient ensuite le problème de la détection de sources kimberlitiques à partir de données d'indicateurs minéraux. Diverses méthodes statistiques ont été testées, la méthode de l'arbre de classification s'avérant la plus probante. Cette méthode répartit les données en classes de distance à partir d'une source, basée sur un modèle construit en utilisant les données au voisinage de kimberlites connues. L'application de cette méthode fait l'objet de deux cas d'étude. Finalement, une technique est mise au point pour évaluer l'efficacité de l'échantillonnage par reconnaissance aérienne. Elle repose sur des calculs de contours polygonaux et de leurs intersections. Par ailleurs, une fois que les indicateurs minéraux les plus signifiants ont été identifiés, la forme de leurs nuages est modélisée. On s'inspire alors du problème de l'aiguille de Buffon pour déterminer quel espacement prendre entre deux lignes de

reconnaissance aérienne, de façon à rencontrer les nuages des indicateurs minéraux avec une probabilité donnée à l'avance.

Chapitre 4: optimisation de l'échantillonnage en phase d'exploration

On cherche à simuler des environnements de dépôts en présence d'un nombre limité, sinon nul de données quantitatives. Bien souvent d'autres données existent en abondance, que le géologue interprète mentalement et utilise lorsqu'il s'agit de définir des stratégies d'échantillonnage. C'est cette perspicacité et cette expérience que nous cherchons à reproduire au moyen d'esquisses au crayon. En chaque point, l'intensité du figuré est proportionnel à la valeur probable de la densité, les zones les plus riches étant représentées en sombre. Il existe une méthode rapide permettant de convertir cette esquisse en simulation; le résultat est approximatif mais bien utile pour optimiser l'échantillonnage en l'absence de toute donnée quantitative. Dès que ces données deviennent disponibles, les méthodes spatiales de simulation géostatistique peuvent s'appliquer. Tout ceci est illustré sur un cas d'étude pour lequel une simulation a été créée. Dans des environnements géologiques complexes, des variogrammes directionnels locaux ainsi que des configurations locales de voisinage de krigeage ont été envisagés dans le but d'améliorer les estimations. A titre d'exemple, on montre l'influence de ces paramètres en appliquant divers scénarios à un petit jeu de données de 5×5.

Chapitre 5: optimisation de l'échantillonnage des placers sous-marins

Des recherches ont été entreprises pour quantifier la perte de précision obtenue en employant des échantillons de taille inférieure à celle d'usage courant. Pour cela, un dépôt marin a été simulé (conditionnellement, à l'aide du processus de Cox), et sur la simulation obtenue plusieurs scénarios d'échantillonnage ont été effectués et comparés. Dans le cas d'étude présenté, il s'avère qu'il serait préférable de poursuivre l'estimation des densités de pierres au moyen d'échantillons de $10m^2$.

Chapitre 6: échantillonnage et estimation des terrils

Dans le but d'échantillonner un terril à Kimberley, on a cherché à développer des méthodes scientifiquement fiables pour optimiser la taille des échantillons et leur espacement. Les lois de Poisson et de Sichel ont été utilisées pour établir la taille minimale des échantillons. Cette taille a été comparée à celle fournie par Pierre Gy avec son calcul d'erreur d'échantillonnage fondamentale. Une autre méthode est aussi proposée pour estimer des blocs par moyenne pondérée en tenant compte de la direction de croissance du terril et de l'angle de talus. Des modèles sont utilisés pour justifier les méthodes proposées. Un rapport de projet (en référence) documente l'implantation de la méthode d'estimation choisie.

Annexes

On y a consigné les bases des méthodes proposées ainsi que quelques compléments.

A) Bénéfice qu'une compagnie peut tirer d'une telle recherche

Les recherches effectuées dans les années 90 ont été considérées comme bénéfiques, mais

jamais évaluées en termes monétaires. Un exemple très simple montre l'apport financier d'un changement de taille d'échantillon.

B) L'aiguille de Buffon

Ce problème correspond exactement à la situation d'une minéralisation en forme de lentilles allongées et d'un échantillonnage en lignes parallèles.

C) Echantillonnage d'une minéralisation très dispersée

Dans les gisements côtiers situés en bordure atlantique de l'Afrique Australe, les diamants sont nichés dans des irrégularités du socle (ravineaux, marmites de géant). Les travaux de recherche sur l'échantillonnage de ce type de minéralisation très dispersée, portant sur la taille, la forme, l'orientation des échantillons et leur disposition spatiale, ont été présentés en 2005 lors de la Second World Conference on Sampling and Blending [41]. L'approche utilisée est applicable à bien des problèmes d'échantillonnage des gisements secondaires.

D) Krigeage par voisinages directionnels

Dans les environnements chenalisés, on peut améliorer l'estimation par krigeage en introduisant des directions locales d'anisotropie correspondant à celles des chenaux. Un exercice théorique a été mené en reprenant les variogrammes observés sur ce type d'environnements et en étudiant le comportement des poids de krigeage obtenus. De façon surprenante, un effet d'écran est constaté. Des résultats contre-intuitifs ont également été observés pour certaines données et certains variogrammes, ce qui suggère de n'utiliser qu'avec prudence les voisinages dynamiques dans les environnements chenalisés.

E) Résumé sur l'erreur déchantillonnage fondamentale de P. Gy

On reprend la description de François-Bongarçon [25] concernant les différents constituants de l'erreur totale d'estimation. L'erreur fondamentale d'échantillonnage est aussi présentée et illustrée à la manière de Royle [58]. On en profite pour donner les valeurs typiques que prennent les paramètres dans les applications.

F) Simulation conditionnelle du processus de Cox

Le processus de Cox [42] est le modèle choisi pour simuler des gisements de diamants conditionnellement. On décrit ce modèle ainsi que la procédure utilisée pour échantillonner les gisements souis-marins.

Chapter 2

Introduction

2.1 Diamond placer deposits

Kimberlite pipes are the primary source of diamonds and were created millions of years ago through upwelling of magmatic material through a volcanic process. The pipes were eroded over time and in Southern Africa, the material was transported via fluvial processes to the coast and then reworked by wave action and moved northwards by the Benguela current. During the transportation process, poor quality diamonds are destroyed and final placers deposits that were formed contain diamonds that return of the highest average revenue per carat in the world. The diamonds are finally deposited in a range of placer types that include fluvial placers, beaches, marine deposits and aeolian deposits.

The secondary deposits in the marine environment were formed by several phases of reworking during transgressive and regressive periods and intermittent sea stillstands. Storm events redistributed the diamonds across wide areas where secondary concentration of diamonds is related to geological features, for example gullies. Periodic flooding of the Orange River system resulted in "pulses" of diamondiferous gravel being introduced into the sea.

A combination of erosional forces and the characteristics (hardness, type of material, slope, etc.) of the bedrock result in so-called trapsites that take the form of potholes and gullies in which the diamonds are concentrated.

The diamond placer deposits found along the South West African Coast of Namibia and South Africa are unique in character and comprise a truly world class deposit.

These deposits are currently exploited onshore (including beaches, rivers and embayments) with large scale mechanical mining equipment. In a marine environment, at depths of around 100 to 140m below sea level, the deposits are exploited by mining ships and closer to shore, deposits at depths to 30m are mined by divers.

The deposits vary in size and through a process of systematic identification of target areas, sampling and evaluation is necessary because selective mining methods are required.

All of these deposits and including the tailings resources (dumps) are classed as secondary deposits as they are not the primary source of diamonds and form the basis for the case studies of this research.

2.2 The importance of reliable sampling

Accurate mineral resource estimation requires good sampling and it is important that this is done correctly to minimise risk and improve decision making during mine planning and production phases. It is therefore important to ascertain the purpose of the sampling and clear objectives should be defined for each programme.

Sampling can have different objectives and different programmes may conflict with one another. By way of an example, consider exploration sampling. Some exploration geologists will plan to sample in a severely biassed way as the objective of the sampling is to prove mineralisation, as well as to obtain an indication of the grade. For mineral resource estimation this sampling is not adequate and for best results an unbiased sample campaign is required, with samples taken across the whole deposit, covering all geological environments to establish a representative grade.

What exactly is a representative sample? This question is deceptively simple and when a few definitions are considered, they appear to differ and this suggests that the concept is complex to define. The definition also changes depending on whether a statistical, geological or mining book is referenced. A number of definitions for sampling are given below[1]. Sampling is:

- the process of taking a small part or quantity of an unknown entity, using various techniques, with the intention of showing, at a level of confidence, what the whole could consist of.
- the creation of a statistically significant subset selected and analysed by an industry accepted method or measured by an appropriate technique to estimate the characteristics of a larger group or population (Jorc Code 2004).
- described in the geostatistical glossary by Olea [51] in two ways:
  (i) in statistics, as the creation of a set of all those observations actually recorded, with the sample a subset of the parent population and
  (ii) in geology, considered as the collection of observations or specimens in a survey.
- defined as the operation of removing a representative part, convenient in size for testing, from a whole of much greater bulk, in such a way that the proportion and distribution of the quality to be measured are, within reasonable limits, the same in both the whole (population) and the part removed (the sample) [2]

For the purposes of sampling for diamonds and for evaluation, the last definition is the most comprehensive and discussed in more detail. It is logical and intuitive to use, but a careful study of the definition highlights the need for a greater understanding and realisation that the process of sampling is complex and the ease of doing it well should not be underestimated.

2.2.1 Components of sampling and the introduction of sample error

The definition of sampling by Covacic combines a number of key elements.

Sampling is defined as the operation$^{(A)}$ *of removing*$^{(B)}$ *a representative part*$^{(C)}$*, convenient in size for testing*$^{(D)}$*, from a whole of much greater bulk*$^{(E)}$*, in such a way that the proportion*

[1] *Guidelines for Mineral Resource Management* by SP Duggan, Internal De Beers Report, April 2008
[2] *Sampling Techniques*, Mining Magazine by D Covacic & J Clarke, October 1990, p270.

and distribution$^{(F)}$ of the quality to be measured are$^{(G)}$, within reasonable limits, the same$^{(H)}$ in both the whole and the part removed$^{(I)}$: the sample

Each of the components making up the definition is briefly discussed, highlighting difficulties and show opportunities for the introduction of sampling error:

Sampling is

$^{(A)}$ **defined as the operation:** For operational reasons in secondary deposits and due to the large volumes required to measure the grade of material, production plants are used to treat some material as "samples". In such cases, it is known that the diamond production plants are not suitable for accurate measurements and there will be a lack of sample integrity.

$^{(B)}$ **of removing:** When holes are drilled or trenches are excavated in unconsolidated material, the material could cave in and result in the contamination of samples. This create difficulty in determining accurate measurements for sample volume and weight.

$^{(C)}$ **a representative part:** What does this mean and how is this measured? For diamond grade measurements, bigger samples are usually better, but at what point have micro structures (the nugget effect) been eliminated? The discrete nature of diamonds require a minimum sample size to be taken, as very small samples are subject to big fluctuations in measurements due to the binomial nature of the diamonds. Due to the discrete nature of diamond occurrence, it can happen that small samples contain no diamonds at all, yet the area is mineralised.

$^{(D)}$ **convenient in size for testing:** Sample sizes when sampling for diamond grade can vary from a few tons of material to as much as 40 000 tons. In some cases, even larger volumes of material (up to 150 000 tons) were treated to establish the grade in tailings mineral resources. Practical limitations in the ability to treat large volumes of material will reduce representivity.

$^{(E)}$ **from a whole of much greater bulk:** This implies that "the whole" is known. In some cases the geological model is difficult to determine and can comprise a combination of multiple smaller geological domains.

$^{(F)}$ **in such a way that the proportion and distribution:** Proportion and distributions imply that statistical models can be used to summarise the data. It must be highlighted that the sample is one realisation of reality and taken from a specific geological environment and is therefore subject to variability in the outcome.

$^{(G)}$ **of the quality to be measured:** Measurement errors can occur when diamonds are lost as part of the metallurgical recovery process.

$^{(H)}$ **within reasonable limits the same:** Reasonable limits and a measure for "the same" are dependent on the level of acceptable risk. Results are never absolutely correct and subject to interpretation. Reasonable limits could also be considered results that are repeatable, with a low probability of statistical variation in the outcome.

$^{(I)}$ **in both the whole and the part removed, the sample:** If samples are not spatially collected across the whole deposit there could be a bias in the results.

Taking cognisance of the components of the sampling process will ensure good quality and reliable samples for use in estimation and simulation exercises.

Once a sample had been taken, it becomes part of the data collection that will remain in existence for the duration of the project. Figure 2.1 shows the process and introduction of sample data along the mineral resource cycle. The figure highlights that initial sampling in a project has the greatest impact, remains in existence the longest and it is critical that this is done correctly. It is at this early stage, with insufficient data for an in-depth sample optimisation, that it is most difficult to design a sample programme.

The discussion above and potential introduction of sample error must not leave the impression that good sampling is impossible to achieve, but rather that care should be taken during the sampling process and planning to ensure the integrity of the samples.

2.3 Sampling challenges

With a constant need for sampling of grade, benefit is derived by having the sampling parameters optimised. This benefit comes in the form of more reliable and representative samples which in turn give rise to improved profitability of the mining operation.

For example, in the deep water marine environment (100 to 140m below sea level), the geological homogeneous zones are fairly large and continuous. Research was undertaken between 1990 and 1996 to establish an optimal sample size and spacing for such an area. The outcome of the research and guidance regarding optimal sampling parameters are still in use today and has benefited the company financially for more than 15 years (see Appendix A).

Currently, more challenging environments with lower grades are targeted for exploration with questions regarding sample size and spacing to achieve accurate grade estimates. In some cases, very little grade information is available and sometimes no information is available at all and methods have to be researched to optimise the sampling effort.

In such cases, assumptions about stationarity cannot always be made, as clear geological boundaries have not yet been established, are too complex to determine based on the available geological data or have no grade information to help with the determination of the boundaries. In these situations, even an approximate result would be helpful in guiding the sampling campaign.

The most pertinent questions remain around the parameterisation of the sampling and the emphasis on the taking of good quality samples.

Some examples are discussed below to show the subtlety and difficulty sometimes encountered to sample, simulate and interpret results accurately.

2.3.1 Examples of challenges encountered while sampling

In the research, reference will often be made to the Mega-drill and Deca-drill. Following are pictures describing the sample tools used for deep water marine sampling.

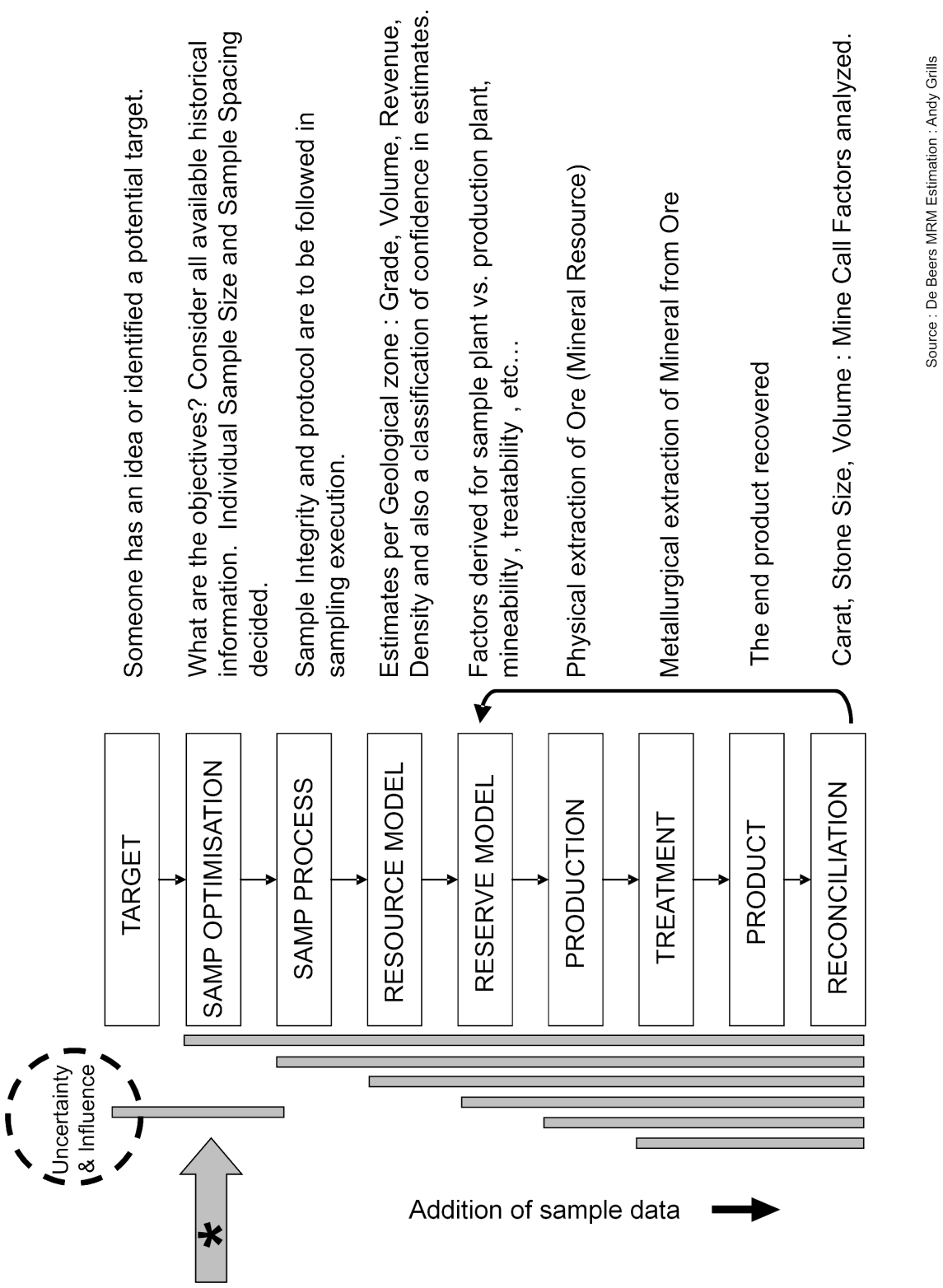


Figure 2.1: The introduction of sample data along the mineral resource cycle

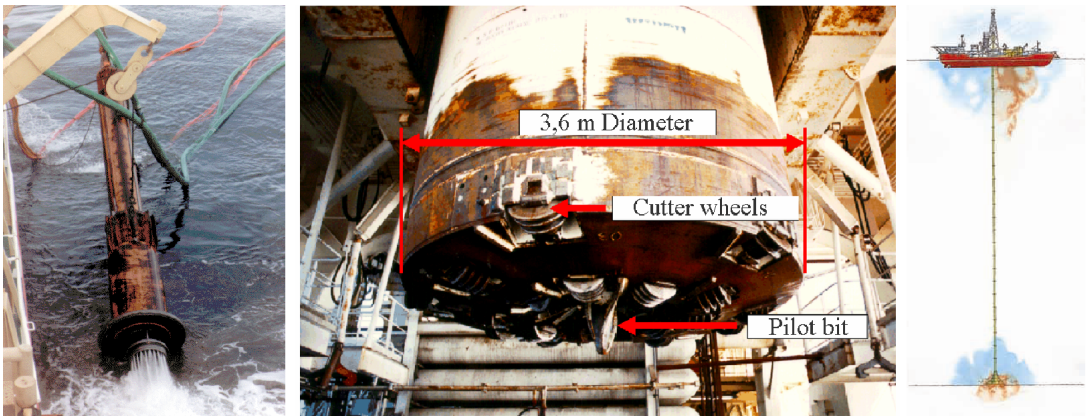


Figure 2.2: Sample tools used in the marine environment: Mega-drill (left) and Deca-drill(right)

Mega-drill: Figure 2.2 (left): This is a percussion tool with size 0.72m^2 , taking a sample by bouncing the tool (within the skirt), until the flow of gravel stops or footwall is reached. The bottom of hole is determined by looking at the type of material going over the screens onboard the ship.

Deca-drill: Figure 2.2 (right): This is a rotary Wirth drill of 10m^2 , showing the cutter wheels and pilot bit. The drill is attached to the ship through a series of pipes which provides the torque to turn the drill and simultaneously act as a pipe through which the sampled material is recovered.

The material obtained from the samples are treated onboard a ship through a treatment plant and diamonds (referred to as stones) are recovered. Once the material is treated it is discarded back into the sea.

Requirement for large samples to recover sufficient numbers of diamonds in very low grade areas

At Sendelingsdrift in Namibia a palaeo-fluvial gravel deposit with very low grade was targeted for sampling. Samples up to 40 000 tons were taken to ensure that sufficient numbers of stones were recovered for use in grade estimation. The number of stones recovered per sample ranged from 121 stones in 37 000 tons of material to 2 944 stones in 21 000 tons of material.

When these samples were planned, a debate was held regarding the merits of many small samples versus a few large samples. The outcome was to use a few large samples. This was determined by using a simplified simulation method, before more sophisticated methods were in use.

Figure 2.3: Sample size and locations at Sendelingsdrift, a deposit along the Orange river between Namibia and South Africa

These samples are 40x40m in size and eight samples were taken to represent the $\pm 8 \times 1$km area. The samples were carefully selected in terms of the number and their positions were determined according to detailed interpretation of the geology.

"Worm holes" using the Deca-drill

Tests were carried out with the Deca-drill to establish the effectiveness of the sample tool with and without a pilot bit. Without the pilot bit, the drill could not penetrate the seabed properly as shown below.

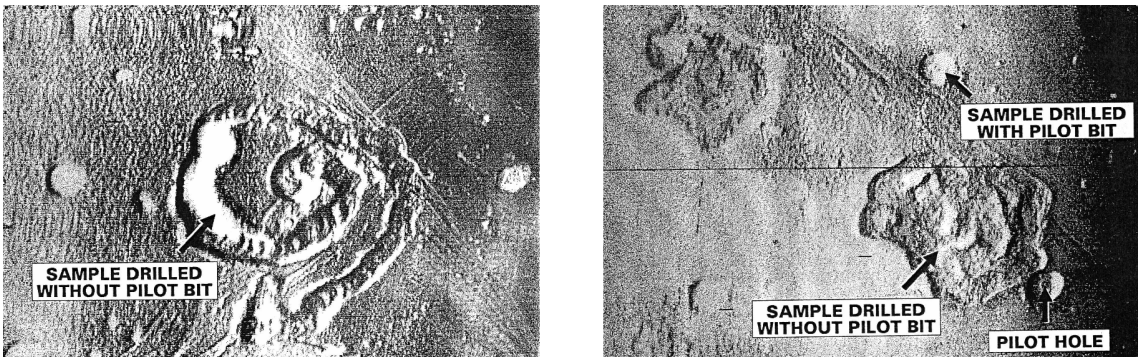


Figure 2.4: Two examples showing samples taken with and without the use of a pilot bit

Large scale non-representative sampling would take place if holes are drilled without a pilot bit. An effective length for the pilot bit had to be established to cater for areas with soft and hard footwall characteristics. A long pilot bit would not be a problem in soft footwall areas, but in hard footwall areas problems with penetration can occur and a shorter bit would be preferred. An optimal length was determined by trial and error.

Over penetration of samples (A SASA example)

In the South African sea areas (SASA), both the Mega- and Deca-drills were used to generate samples for estimation. As the footwall type changes from harder clay footwall to softer sandy footwall, the sample efficiency changes, sometimes over very short distances. The softer footwall causes sample contamination by material being sucked in from the side of the sample tool. Different skirt configurations were tested in an attempt to reduce the effect of this problem. Figure 2.5 shows the results of this non-representative sampling of Mega-drill samples where there is a change from a clay to a sandy footwall area.

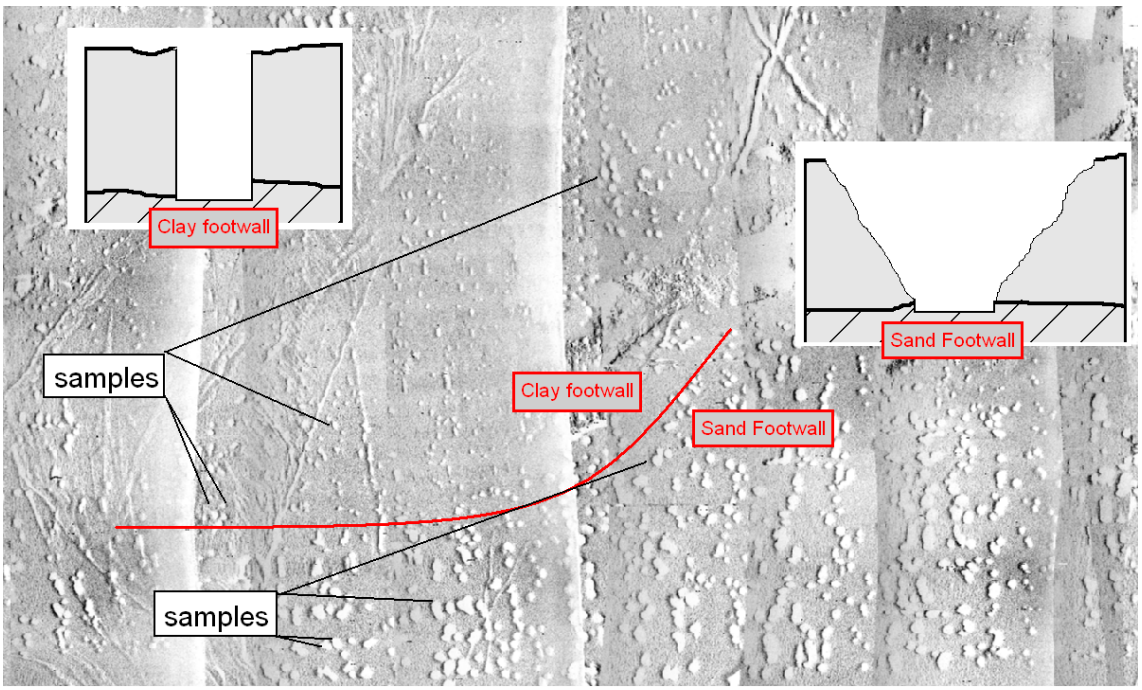


Figure 2.5: Effect of footwall on sample integrity: Variation from clay to sandy footwall

Sample integrity is some areas is good, an example of which is shown in Figure 2.6.

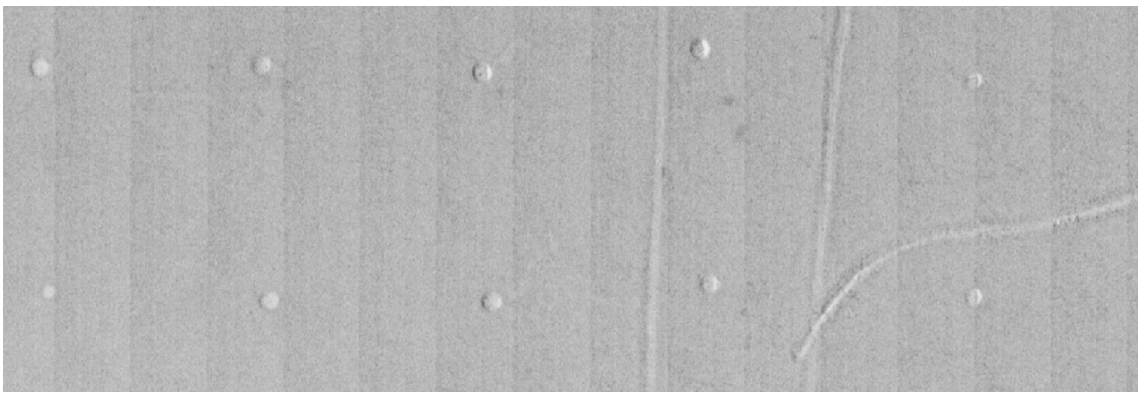


Figure 2.6: High integrity $10m^2$ samples (although it cannot show if samples penetrated onto footwall)

The extent to which samples are contaminated is unfortunately not measured at all the sample locations and is difficult to quantify in the estimation process.

Dealing with a bent bit

During the initial testing period of the Deca-drill, different lengths of pilot bits were tested. At a stage it was observed that the pilot bit was bent during the process of taking samples. The offset at the tip of the pilot bit was 400mm. With the drill's rotation and offset caused by the bent pilot bit, a larger sample was taken. Although 400mm does not seem a lot across a diameter of 3.6m, the change in area of such a sample could be substantial.

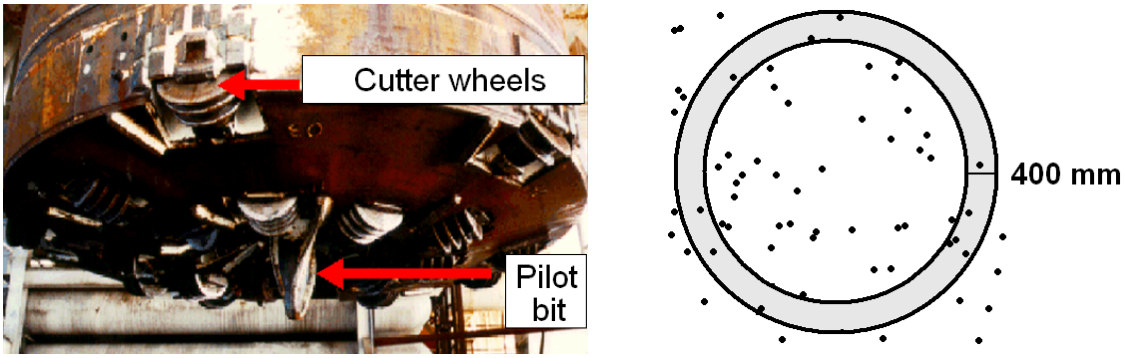


Figure 2.7: Effect of the bent pilot bit

The difference in area between a 3.6m diameter sample and a 4.4m diameter sample is an increase from 10.18m^2 to 15.2m^2 . The bias is an increase in sample size of 49%. This problem was identified and rectified immediately.

Even without problems concerning the pilot bit, the rotary action of the drill causes a similar bias in the sample results.

A more common problem is with material coming in from the side of the sample tool, as holes are drilled in very soft sediment. The ability to clean out the sample hole at the bottom, as the tool is extracted also causes an inefficiency in the sample measurement.

These problems are continuously addressed through new designs of the channel through which the material is extracted and skirts for the sample tool in an attempt to improve the engineering efficiency of the sampling and mining systems.

A shortened pilot bit (as in the figure above) is currently used.

Effect of "roll-over" of stones

While sampling is in progress, diamond simulants are placed into the stream of material being treated, to test the efficiency of the metallurgical recovery process. A known number of simulants are introduced into the system and being of the same density and luminescence as diamonds, are recovered along with the diamonds. An estimate can thus be made of the efficiency of the recovery system based on the number of simulants recovered from the samples into which simulants were introduced, as well as the following samples, thus measuring the extent of the "roll-over" of stones.

For one such a sampling exercise, it was determined that there was significant roll-over of

stones present in the data. An attempt at correcting the data was made as it was required for use in a simulation exercise.

The date and time each sample was recovered is used to sort the data chronologically and in combination with the monitoring of the metallurgical efficiency, the stones were "rolled back" into their respective samples.

The method considered that if O_n is the observed stones in a sample n and X_n is the actual number that should have been present, each observation of stones consists on average of 85% recovered stones, 12% of stones rolled over from the previous sample and 3% from the sample before that (based on simulant recovery), i.e.

$$O_n = 0.85X_n + 0.12X_{n-1} + 0.03X_{n-2}$$

The actual number of stones per sample is X_n which can be expressed as:

$$X_n = \frac{O_n - 0.12X_{n-1} - 0.03X_{n-2}}{0.85} \tag{2.1}$$

Similarly,

$$X_{n-1} = \frac{O_{n-1} - 0.12X_{n-2} - 0.03X_{n-3}}{0.85} \tag{2.2}$$

Substituting equation 2.2 into 2.1, eliminating terms that become insignificant and after simplification, X_n is expressed as:

$$X_n \approx 1.18\ O_n - 0.17\ O_{n-1}$$

The model can only be considered as an approximate correction for the true underlying data and is an undesirable side effect of sample inefficiency.

Bias between sample and production data statistics

During a simulation exercise undertaken to investigate the possibility of using a larger sample size to sample the Atlantic I area by Prins [53, 54], the only data available to determine the simulation parameters were from the Mega-drill and production data. A statistical comparison between the two datasets showed significant differences in the means. The production data had a higher mean than the sample data. This is attributable to the fact that the sample data is evenly spaced across the area but the production data was from a smaller, high grade targeted area, selected to return a profit when mined. The production data representing the low grade part of the distribution is left unmined and is therefore not reflected in, thus biasing, the production statistics upwards.

The production data were considered more reliable, but the parameters required for the simulation were required at point support. Deconvoluting from production support (50x50m blocks) to point support (2.16m^2), a size ratio of 1157:1 could not be done successfully.

The Mega-drill sample data was the only other source of information, but lacked the reliability to determine the statistics for creating a suitable simulation. This caused difficulty in determining the parameters for the simulation model.

Before the parameters of the simulation were finalised, attempts were made to establish suitable parameters on point support to ensure that, for the verification process, the 50x50m blocks from

the simulation would be comparable with the 50x50m blocks and statistics of the production data. Graphically, a few iterations were required to adjust and find the appropriate parameters:

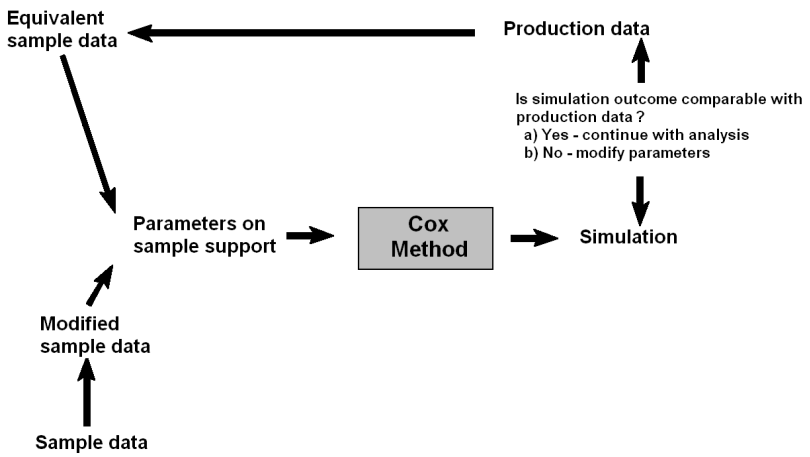


Figure 2.8: Iteration of parameters on point support to ensure compatibility between production data and 50x50m simulated blocks

This process was carried out successfully and the simulation parameters were determined.

Complexity of marine geological environments

Some geological terrains in the marine environment can be of a complex nature. The figure below shows the complexity of a regional geological interpretation in which each colour represents a different type of geological facies, based on the interpretation of remotely collected data.

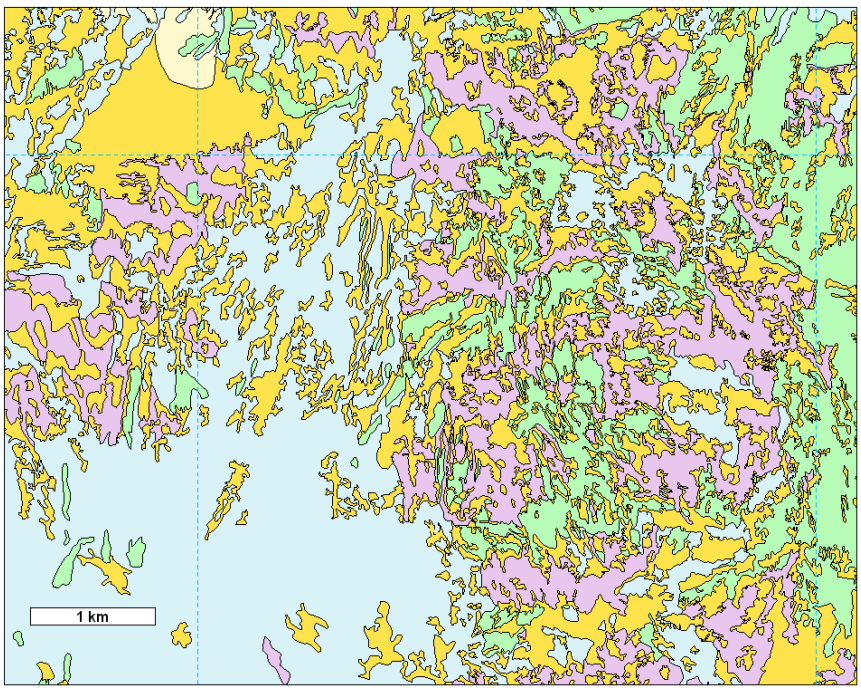


Figure 2.9: Complex system of entrapment features and geological facies

The area is not a large continuous stationary environment and would be challenging to sample and evaluate. The directional nature of the geology introduces the possibility of using locally varying angles for kriging (see appendix D).

Mixing different sample support sizes

Different sample sizes and configurations are often encountered in data and have to be dealt with in simulation and estimation procedures.

An example showing sample pits and trenches sampled on the outskirts of Oranjemund (Namibia) is shown in figure 2.10. These different sample campaigns are the result of the development of new ways of sampling and changing economics over time.

Figure 2.10: Examples of pit and trench sampling

Similarity can be drawn to sample optimisation studies required for marine environments where mixed sample support sizes (0.72m^2 , 10m^2 and 50x50m block data) are available. Due to the characteristics of particulate data, methods adapted to take cognisance of the different sample supports are required.

The impact of cemented horizons on measurement of grade

The depositional environment in some marine deposits has resulted in cementation, locking the diamonds in hard material. The grade of the cemented material is the same as the surrounding material, but it is however not always possible to sample and is difficult to treat and recover all the diamonds from.

Onshore at Elizabeth Bay mine (near Lüderitz, Namibia), extensive areas of the deposit comprise cemented material.

Sampling activities at Elizabeth Bay is shown in figure 2.11.

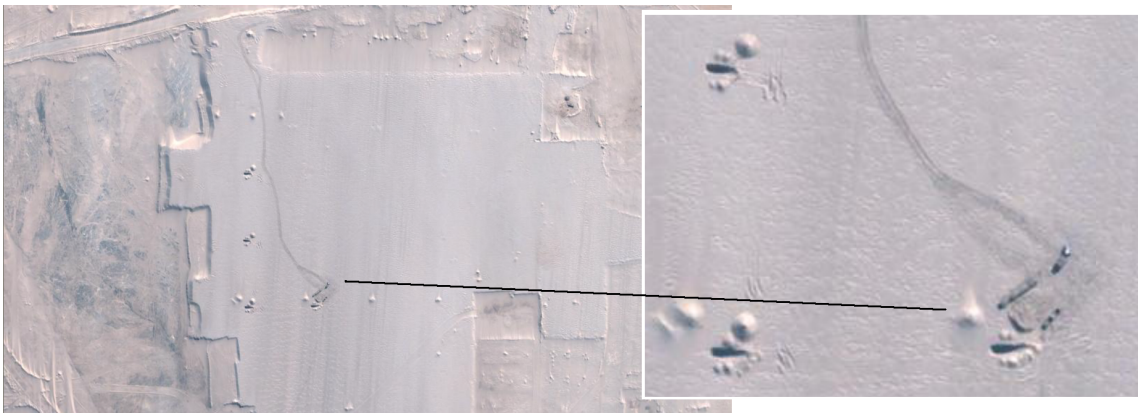


Figure 2.11: Elizabeth Bay Namibia: Sampling activity

Conventional sample tools did not always penetrate the cemented material, resulting in inconsistent thickness measurements. The cemented areas are variable in size and in many instances the thickness of the deposit was assumed without taking cognisance of stratigraphic horizons.

This combination of circumstances affects the measurements of grade and thickness and have an impact on the accuracy of the estimates.

Measuring grade in highly variable erosional environments

In beach deposits, gullies and potholes contain a high percentage of diamonds, but occur with spatial randomness that is impossible to predict. In some cases, the sampling equipment is aggressive enough to drill into the footwall, but if this is not the case, the representivity of the samples is questioned, especially with the possibility that the sampling will show a lower than expected grade. An example of well established entrapment features is shown in Figure 2.12.

Figure 2.12: Random occurrence of gullies and potholes

For more marginal projects, sample measurements are required to determine the degree of mineralisation contained within the gullies. As part of internal De Beers consultancy [3] and an ongoing project report[4], research is under way to express the probabilistic outcome of gully penetration by varying the size of sample tool, to express the accuracy of grade determination in channelled areas.

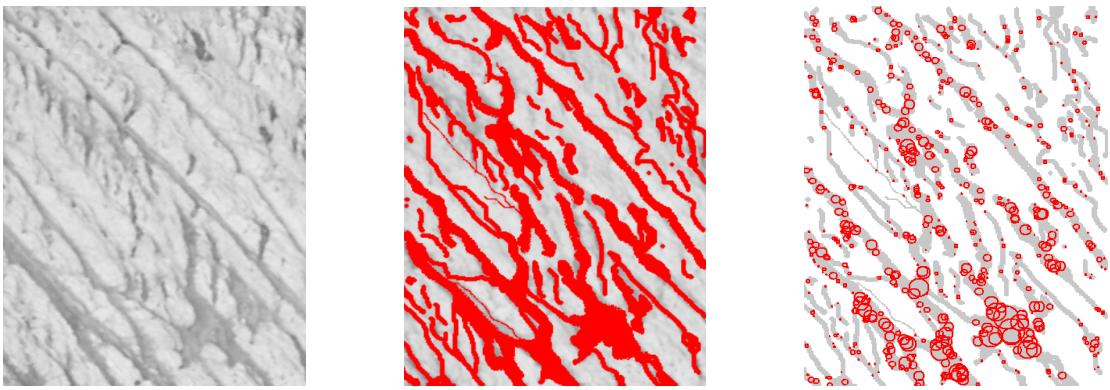


Figure 2.13: Studying the ability of different sample sizes to penetrate gullies. From left to right: Original image, interpreted image and subset of sample locations and size that penetrated gullies

For marginal economic projects, the ability to express the grade estimates relative to the ability to penetrate gullies becomes important.

In summary:

The sampling and mining of secondary deposits require skill and constant optimisation of the sampling, simulation, estimation and mining parameters. The examples shown gives an appreciation of how varied the challenges and environments are in which sampling and estimation is undertaken.

Improvement is an ongoing process and is continuously driven in the areas of sample optimisation and accuracy of estimates and in researching this, different topics were identified.

2.4 Research undertaken

Topics related to the sampling and simulation of secondary deposits form the basis for two chapters of the research and two other topics have been researched related to the sampling and estimation of tailings resources(dumps) and the analysis of data related to kimberlitic indicator minerals and exploration target identification.

The chapters of this thesis are not related to one another, but have been sequenced to firstly cover exploration, then sample optimisation in secondary deposits and lastly the sampling and evaluation of tailings resources. A short summary for each topic is given.

[3] *Determining largest circular sample size across a network covering a gullied area* by CF Prins, Internal Deers report, March 2009

[4] *Investigating gully penetration using different sample sizes* by J Jacobs, Ongoing M.Sc project report, WITS University, Ongoing research: September 2009

2.4.1 Clustering of kimberlites and identification of exploration targets through analysis of indicator mineral data

The exploration for kimberlites is a high priority for De Beers and any method by which success can be achieved at less cost or more efficiently will be of benefit to the company.

Data were made available by the De Beers Exploration departments with a request to research any (geo)statistical methods that could be applied to assist in finding kimberlites. Known kimberlite occurrences on the Southern African and Canadian craton areas are present in the data and are used in model building and case studies.

Research:

The research comprised different parts:

Firstly: A plot of the kimberlite locations on the Southern African craton showed the possibility of clustering. This clustering was a topic of research for another internal De Beers study by Barnett [3], but geostatistical methods and cluster analysis as implemented in this thesis did not form part of Barnett's study. This first part of the research is aimed at confirming the visual observation of clustering and then to determine the areal extent of an average cluster size.

Knowledge related to the size of clusters is required for:

a) The application of an exploration license that gives the company an opportunity to explore a specific area for a certain amount of time and depending on the country, followed by a process of handing back percentages of the area on a "use-it-or-loose-it" basis. Knowing the size and behaviour of kimberlite anomalies and cluster sizes will result in an informed decision.

b) To specify the areal extent once an anomaly has been identified and target the correctly sized area for airborne geophysics. If a kimberlite anomaly is identified and it is believed to be of interest, it will be beneficial to retain and explore the correctly sized area around the anomaly. This will maximise the probability of finding the pipe's location and surrounding kimberlites.

Secondly: Large areas of Southern Africa and Canada are covered by Indicator Mineral (IM) measurements. The indicator minerals occur in much higher proportions than diamonds and are easier to observe in and around a kimberlite location. The composition of indicator mineral profiles are dependent on the type of kimberlite and migration behaviour of the particles, given that they have different properties, for example size, hardness and density. Indicator minerals migrate in varying proportions up to many kilometres from a kimberlite source and when observed in relative high frequencies, generally indicate the proximity of a kimberlite.

It is challenging to find a method to analyse this data, as many factors play a role in the occurrence and migration of indicator minerals and geological knowledge, combined with the indicator mineral results provide the best opportunity to identify and target anomalies. The area over which the analysis was undertaken includes different geological regions and the indicator minerals occur in different proportions depending on the type of kimberlite and are subject to different weathering and erosional forces.

Statistical data mining techniques were investigated to establish a suitable method for analysing the data and to identify potential target areas.

2.4.2 Simulating complex placer environments using a hand sketch

The continuous sampling of secondary deposits is required because most placer deposits require selective mining. Where an abundance of data is present, spatial simulation techniques such as described in appendix F can be applied to optimise sampling campaigns.

There are situations where a sample optimisation study is required, but no data is available, for example a new target that was previously not sampled.

Such areas may have probe samples that were required for the geological interpretation and to outline the target. With this information and making use of the interpretation by the geologist, a sample program will be planned.

It is beneficial for an optimisation study to be undertaken to ensure the correct sample size and spacing prior to the first sample being taken. A method is discussed whereby the available information is captured and the geologist's interpretation included in the model for analysis and used to quantify the effect of different sample strategies.

It should be understood that the model is only as good as the geologist's interpretation and that the analysis can only provide a relative answer, accurate to an order of magnitude. In other words, this answer will be closely linked to the experience and quality of geological interpretation. Once the initial results of grade sampling become available the geological interpretation and optimisation process can be revisited through the use of spatial simulation models.

A method is required to facilitate these early decisions about sampling when the only available data is the geology. One method of capturing the mineralisation model is by making use of a sketch of the deposit using a pencil. The sketching process is carried out by the geologist and is related to mineralisation where darker areas represent higher grades and lighter areas represent lower grades.

The quality and precision of the sketch will determine the validity of the simulation, and although this method is not scientifically sound, no other method could be found for guiding sample strategy at an early stage of exploration.

Consider the example shown in Figure 2.14: Gravel was observed in the deposit after exploratory holes were drilled using a hand auger drill. The geologist is able to construct a model that includes thickness and gravel quality with some idea of diamond potential. The challenge is to decide how to sample the deposit; utilise trenches or a regular grid using a grab tool.

Figure 2.14 shows a pencil sketch of the diamond potential made by the geologist where darker areas represent higher concentration of diamonds.

The geologist's experience, geological interpretation skills and visual observations are applied to draw the sketch. This is considered the capturing of "soft" information, cannot be quantified, yet form an integral part of the interpretation.

Assuming that the only information available for a new target is soft information, a method was researched that by capturing the geologist's interpretation using a pencil sketch, which is statistically modified, a rudimentary simulation can be created. This can be used in the planning of the first sample programme.

Figure 2.14: (Left) Area targeted for sampling and (right) geologist's interpretation of greater diamond potential. The dashed line outlines the target area.

2.4.3 Sample optimisation in marine placer environments

In the mid 1990's research was undertaken by Kleingeld, Prins & Duggan [20, 53, 54] which determined that a $10m^2$ sample was the optimal sample size to use for sampling the Atlantic I marine deposit that would provide block results (50x50m blocks) at an acceptable level of accuracy. This research was carried out using an unconditional Cox process based on non-representative $0.72m^2$ samples, clustered in groups of three making up $2.16m^2$, and 50x50m blocks of production data. Parameters were determined for use in the simulation process. A scaling of parameters to align the sample statistics with the production statistics were required to determine the parameters for the simulation. The exercise was completed, the sample tool size determined and the tool was built and is currently still in use.

In time, the $10m^2$ sample data has become available and has proved to be successful for use in estimation. In an attempt to reduce the cost of sampling and the possibility of targeting different types of geological terrains, there was a requirement to re-examine the outcome of the initial study and test the possibility of implementing a smaller sample tool.

Sample optimisation using a conditional Cox process

The data obtained from the use of a $10m^2$ tool is more reliable than the original data obtained from the smaller tool ($2.16m^2$). Research was done to test the possibility of reducing the $10m^2$ sample tool to a smaller tool. It has been argued that the smaller tool will negate the loss of accuracy in the estimates by producing better and more reliable sample measurements. This investigation aimed to:

a) Use the $10m^2$ data to confirm the results of the previous 1990 simulation project and

b) To test the possibility of using a smaller sample tool in an attempt to enable the deployment of a sample system from a smaller sampling vessel that is unable to accommodate the weight and engineering requirements needed for a rotary type $10m^2$ Wirth tool.

This research aimed to examine different sampling scenarios utilising a conditional Cox simulation method.

2.4.4 The sampling and estimation of tailings mineral resources

Tailings mineral resources (dumps) consists of waste material that has been treated in a diamond recovery facility. Since diamonds are discrete particles of varying sizes, the extraction process involves more than one phase; crushing the material smaller to reduce size fractions from which the diamonds are recovered. Although diamond is very hard, it is also brittle and can be broken in the crushing stage, if not done with care.

Some of the diamonds are not recovered during the recovery process and discarded onto tailings dumps, their loss caused by phased metallurgical recovery methods, diamond lock-up, volume of material treated or minimum diamond size recovered.

The CTP tailings mineral resource in Kimberley is considered a big resource of 120 million tons of material, created over 100 years of mining. Methods were required to find a suitable sampling and estimation method to determine the diamond content of this resource.

For the sampling, suitable methods were required to determine a representative sample size and spacing. A unique requirement was the need for a local block estimate on a grid, for purposes of mine planning.

The sampling and estimation had to produce accurate results as the economic cut-off is close to the average grade of the resource.

2.5 Organisation of the thesis and summary of methods used

The thesis is laid out as follows:

Firstly, the problem of exploring new kimberlite targets are addressed, then techniques are shown to simulate with sparse or absent grade data. A specific marine sample optimisation problem is then researched, with lastly the sampling and estimation of tailings resources.

A summary of headings and technical content for each of the chapters is included.

Chapters 1 and 2: Introduction

Introductory chapters describing background to the research and information related to secondary deposits.

Chapter 3: Optimisation of target areas and analysis of results for kimberlite exploration

Firstly: Kimberlites appear to be located in clusters of kimberlites. Ripley's K function is used to show that they indeed do occur in clusters. Using a nearest neighbour distance calculation, variograms on indicators, Nearest Neighbour and Gabriel graphs, the breakup

characteristics and size of kimberlite clusters are determined. The methods were applied to case studies and results documented.

Secondly: Indicator mineral data were analysed with a variety of statistical methods and the classification tree method was identified as the model of choice. This method classifies indicator mineral data into distance classes from a kimberlite, based on a model built using the data surrounding known kimberlites. This was applied to two case studies.

Finally:

a) A brief description is given on using polygon intersection to express the efficiency of flight lines strategies, and

b) Once the most relevant indicator mineral variables have been identified as part of the classification tree analysis, spatial models are build to show the shape of the indicator mineral cloud. The Buffon needle problem is demonstrated and applied as a method of establishing the spacing for exploration lines, depending on the size of the main axis of the indicator mineral train and required probability of intersection.

Case studies were analysed and results shown.

Chapter 4: Sample optimisation for grade when doing exploration

The simulation of environments with limited or no data is researched. Often there may be a lack of information for grade, but there exists an abundance of other data that the geologist uses and interprets mentally when designing sample strategy. It is this insight and experience that is captured using a pencil sketch, through the introduction of the concept of "soft" data. The intensity of the drawing is proportional to the expected grade. Darker shaded areas represent higher grade. This is a quick method of simulation and will give an approximate result, but is helpful when sample optimisation is required when no data is available. Once the first data is available, spatial geostatistical simulation methods can be used to optimise future sample campaigns.

A case study is shown for which a simulation was created.

For complex geological environments, research was done on the potential improvement of estimates using locally varying neighbourhoods and anisotropic variogram structures when kriging is done. A 5x5 data configuration was analysed with different neighbourhood and variogram scenarios and results shown.

Chapter 5: Sample optimisation in marine placer environments

Research was done to investigate the possibility of implementing a smaller sample tool than was previously determined. A conditional Cox process was used to create a marine deposit that was sampled and evaluated with various sampling scenarios. A case study is shown for which it was found that the $10m^2$ sample size should be retained as the most optimal single sample for use in grade estimation.

Chapter 6: Sampling and estimation of diamond grade in tailings resources

Methods were researched for a tailings resource in Kimberley to optimise a sample campaign and to provide assurance based on a scientific approach to determine sample size and spacing. The Poisson and Sichel distributions were applied to establish minimum sample sizes and this was tested against the fundamental sample error as developed by Pierre Gy.

A method is proposed for localised block estimates using a weighted average, taking into consideration the directional growth of the dump and angle of repose of the material.

Models are documented demonstrating the methods proposed. A project report (as reference) documents the outcome of the implementation of the mineral resource using this method.

The Appendices

Additional information referred to in the thesis and background to methods used in the chapters are provided in the appendices:

A) Expressing the financial value that research adds to a company

The research done in the mid 1990's was considered beneficial, but was never expressed in monetary terms. A very simplistic attempt is documented to express how a change in sample size can be of financial benefit.

B) Documenting Buffons needle problem

Buffons's needle problem is documented, with relevance to the determination of sample line spacing, when mineralisation occur as long thin lenses.

C) Sampling in highly dispersed environments

Diamonds occur in pockets and gullies on the west coast of Southern Africa. With continuous improvement required for sample parameters, research on the sampling challenges in highly dispersed types of mineralisation was done and presented at the Second World Conference on Sampling and Blending (2005) [41], discussing sample spacing, shape and orientation. This approach is relevant to sampling challenges in secondary deposits.

D) Kriging with directional neighbourhoods

In channelled environments, estimates can be improved by restricting the search ellipse and introducing local directional anisotropy in the kriging, along the direction of the channels. A theoretical exercise using typical variograms observed in the marine environment was undertaken. The behaviour of the kriging weights was studied and a surprising screening effect was observed. A counter-intuitive result was seen for certain data and variogram configurations which showed that a dynamic neighbourhood for data and anisotropic variography should be used with care when estimates are done in channelled environments.

E) Summary of P. Gy's fundamental sampling error

A description is given of the components making up the Total Estimation Error, as described by D. François-Bongarçon [25]. The fundamental sampling error, as described and implemented by Royle [59] is documented with a summary of typical parameter values used for applications.

F) Summary of the conditional Cox process

The conditional Cox process [43] is the model of choice for the simulation of diamond deposits. A summary is given describing the model as implemented in the research on sample optimisation to determine sample size in marine deposits.

2.6 Comments

Where applicable, the outcome and application of the research are demonstrated with figures and case studies.

Although the case studies are based on diamond deposits, the methodologies can also be applied to other mineral commodities.

(Deliberately left blank)

Chapter 3

Optimisation of target areas and analysis of results for kimberlite exploration

3.1 Introduction

There are about 5500 known kimberlites in the world, 900 of which are diamondiferous and only 50 that are of economical interest. This suggests that the probability of finding a kimberlite that can be developed into a mine, is very small.

Any method, improvement in technology or strategy that will reduce the discovery time or associated costs is of benefit to an exploration company. Savings made from innovation can be used to target other areas.

Many kimberlites are discovered each year, but most are non-diamondiferous or of little economical interest. As much as it is of interest to find a new kimberlite pipe, it is just as important to have an efficient "walk-away" strategy, especially when a minimum grade threshold seems unlikely to be exceeded.

The focus of this chapter is on methods to help identify new kimberlite anomalies.

Due to the sensitivity of exploration data, limited data was made available for the research, but it was sufficient to identify three main topics and to prove the applicability of the techniques.

3.2 Research Objectives

Research was done to identify possible (geo)statistical methods that could be applied to facilitate the discovery of new kimberlites by accelerating or improving parts of the exploration process.

The main topics identified are:

Investigating the size of kimberlite clusters With knowledge of the expected size of kimberlite clusters, an appropriate area for airborne geophysics coverage can be determined. This can normally only occur once an anomaly has been identified. Knowing the target size can also be used as a guide for licensing application, to ensure areal coverage of the kimberlite anomaly and other possible adjacent kimberlites.

For the case study on the Southern Africa data, the expected size of kimberlite clusters will be determined.

Identifying kimberlite targets through analysis of Indicator Minerals results Soil samples have been collected over many years in large areas of Canada and Southern Africa and analysed for indicator minerals. The statement has been made: "There is potential for discovering a mine in your data" and thus (geo)statistical methods were researched to identify target areas, taking cognisance of the accuracy of historical records. Case studies will be shown where potential areas are identified for comparison with the already identified targets and any new targets will be given attention through focussed exploration. For the identified areas, the indicator mineral data will be rechecked more thoroughly in an attempt to improve the quality of the data and geological assessment will be made of the anomalies.

Sample line spacing A method to establish optimal sample line spacing is documented and will be related to some examples.

In summary, objectives of this chapter are to:

- Find appropriate clustering methods to identify the clusters, then to determine the cluster sizes
- Find and apply multi-variate method(s) to analyse the indicator mineral data to identify potential kimberlite anomalies and
- Determine methods for optimising airborne flight lines and optimal line spacing to intersect indicator mineral trains

3.3 An analysis of kimberlite clustering in Southern Africa

There are many theories [13, 58, 61, 62] explaining the occurrences and localities of kimberlites, of which two primary arguments are:

- The understanding that the kimberlite distribution is controlled by geological structures underlying the earth's crust (either deep crust and/or near surface structures) covered by younger rocks blanketing the kimberlites and
- Theories linking the occurrences of kimberlites with a relationship to hotspot tracks in the earth's mantle.

The kimberlites occur as either isolated occurrences, occurrences in clusters of kimberlites or as multiple intrusions inside previously formed kimberlites.

Although the process of kimberlite formation is still ongoing on a geological timescale, the current occurrences are relevant as a frozen stochastic process with spatial characteristics.

Sample locations for two geochemically different (chemistry, age and petrography) types of kimberlite occurrences (K1 & K2) were made available by De Beers Exploration and another dataset with kimberlite locations was prepared from a digitised map and public domain sources. These datasets will be analysed as case studies to research clustering methods and to determine the size of kimberlite clusters.

The datasets describe:

K1 kimberlites These kimberlite occurrences can mainly be described as pipes, but also with some dykes present. Examples of such pipes being mined are Orapa, Jwaneng, Kimberley, Jagersfontein, Letseng, Koffiefontein, Premier and Venetia. In South Africa no K1 dykes are mined.

K2 kimberlites These kimberlite occurrences can mostly be described as dykes or fissures, with the occasional kimberlite occurring as a pipe. Examples of such occurrences mined are Swart Ruggens, Bellsbank, Robert Victor and Star mine at Theunissen. K2 exceptions that occur as kimberlites pipes that are mined are Finsch Mine and Voorspoed mine.

Southern African craton data These kimberlite locations were obtained from a dataset containing (X,Y) locations, but without geological categorisation. This data were compared and expanded by making use of digitised locations from different public domain maps showing kimberlite locations and prepared as (X,Y) coordinates.

A location plot of kimberlites (figure 3.1) on the Southern African craton area shows that they appear to occur in clusters of different sizes. Furthermore, in relation to the geological factors influencing the creation of the clusters, they also appear to occur in lines of clusters.

One approach to the clustering of the kimberlites is to regard the dataset as a single large cluster and systematically split the cluster up into smaller clusters. By repeatedly doing this an indication can be obtained of the sizes of the clusters formed. Once the process stabilises and at an appropriate scale, it will enable the target clusters to be sized with a good probability of covering the extent of the kimberlite occurrences. This will minimise the areal extent of the exploration area for which yearly license fees are paid. The following paragraphs will clarify this process through a description of the techniques and a case study.

Following, the spatial occurrences of kimberlites, irrespective of the geochemical type, is firstly examined, then the K1 and K2 occurrences are analysed as two geologically different types of occurrences. The locations of kimberlite occurrences in the Southern African data is shown in figure 3.1 and an extract from the K1/K2 dataset is plotted in Figure 3.2.

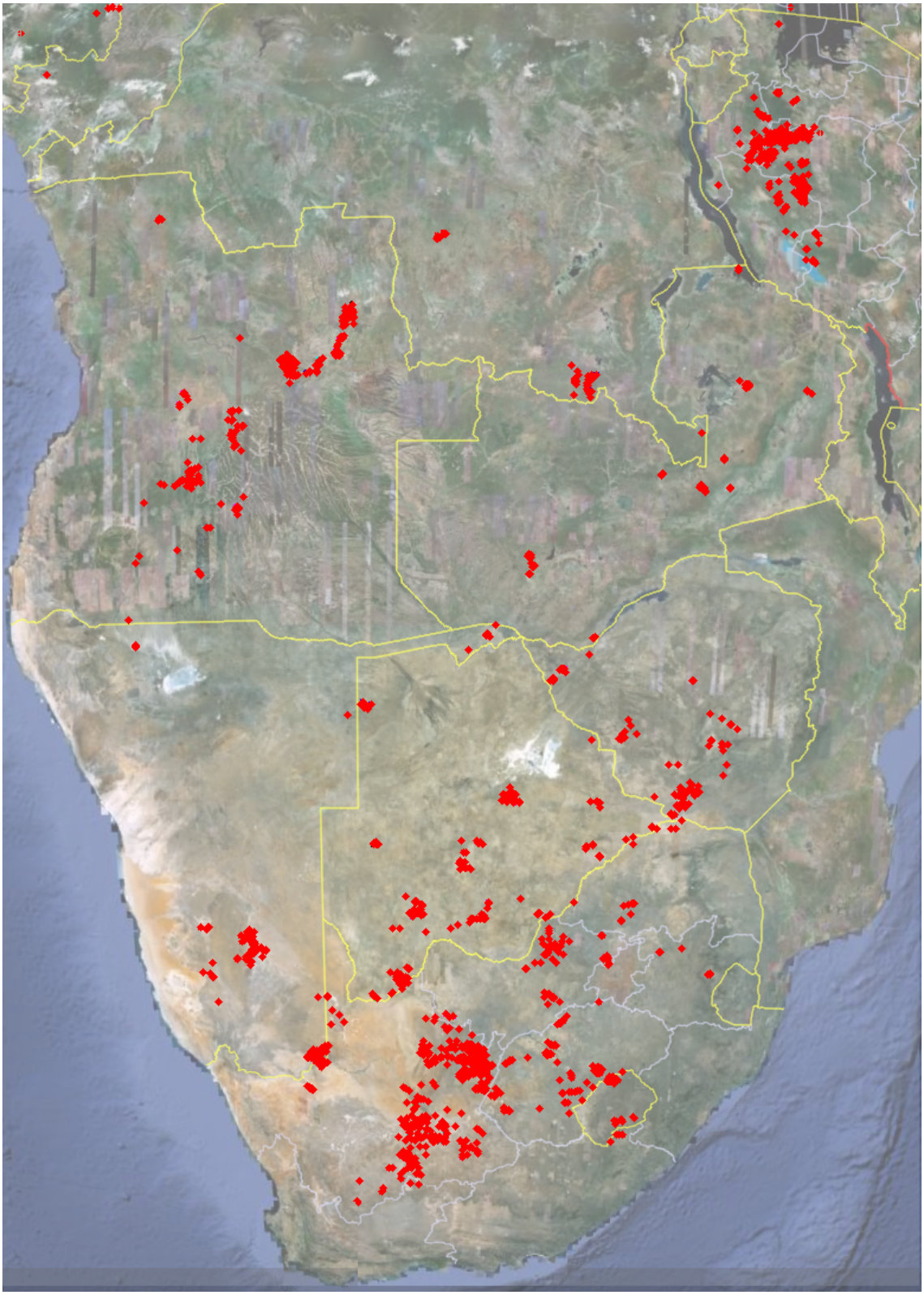


Figure 3.1: Approximate locations of 1929 kimberlites in Southern Africa

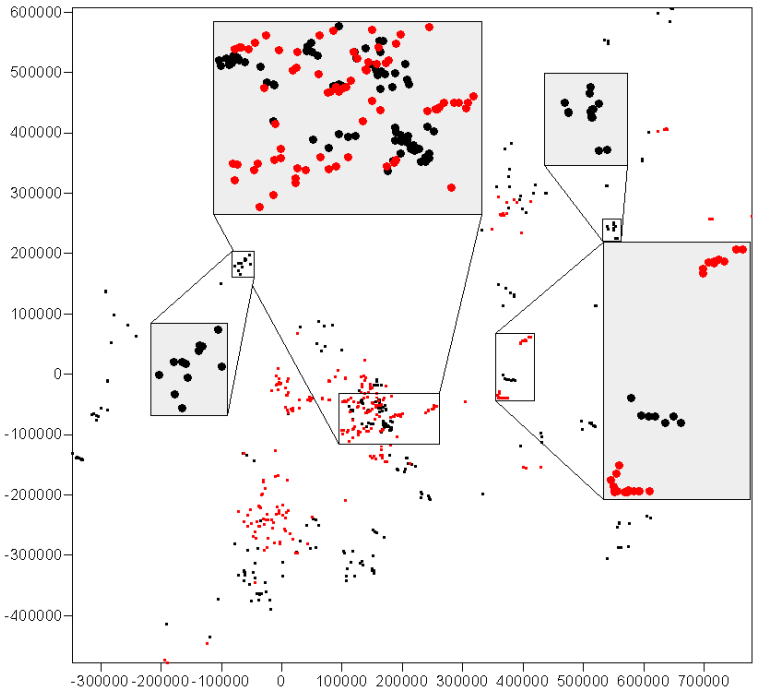


Figure 3.2: Locations of the 310 K1 (black) and 263 K2 (red) type kimberlites showing clustering

3.3.1 Ripley's K function

The clustering of the kimberlites appears apparent by studying the location plots but needs to be confirmed. Ripley's K-function is used to analyse the spatial pattern of point data. It can show global spatial clustering in data by comparing the observed pattern with that generated by a homogeneous Poisson process. Ripley's K function plot as described by Goreuad [31] and Dixon [18] is used.

Following the Goreuad and Dixon's application of Ripley's K function, define the density of kimberlites as

$$\hat{\lambda} = \frac{N}{S}$$

where N is the number of points in an area of size S. Then Ripley's K function considering a distance r is:

$$\hat{K}(r) = \frac{1}{\hat{\lambda}} \frac{1}{N} \sum_{j=1}^{N} \sum_{i \neq j} k_{ij}$$

where $k_{ij} = 1$ if the distance between points i and j is $\leq r$;
and $k_{ij} = 0$ if the distance between points i and j is $> r$.

If the points are spatially homogeneous, $\hat{K}(r) \approx \pi r^2 \quad \forall \quad r$ considered. Rearranging the terms, a plot can be made of $\sqrt{\hat{K}(r)/\pi}$ against r as follows:

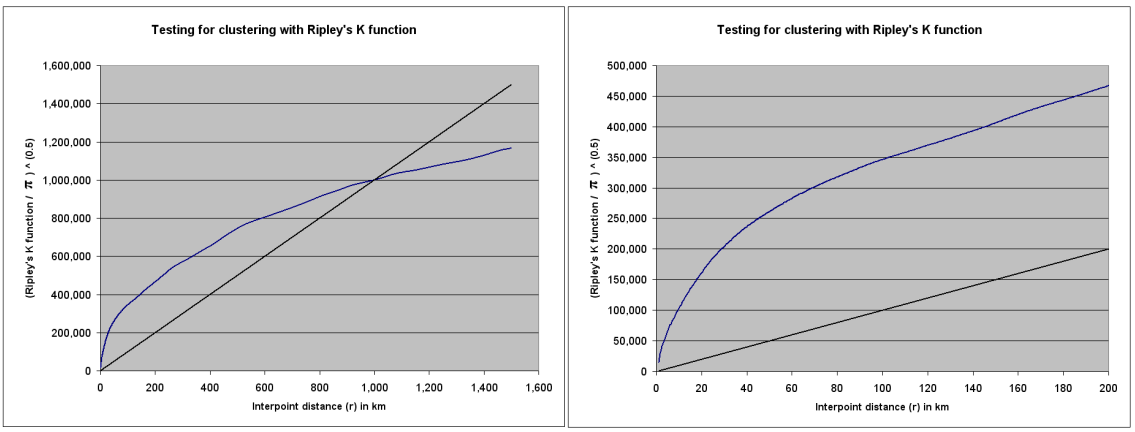


Figure 3.3: Ripley's K-function results for Southern Africa craton data for distances up to 1600km and more detail for distances up to 200km

The deviation from the diagonal indicates clustering up to a scale of 1 000km, with the strongest change in deviation happening up to radii of ±100km.

If $\sqrt{\hat{K}(r)/\pi}$ over distances less than 100km is more carefully looked at and assessing the rate of change in Figure 3.3 (right), changes in the rate is observed up to a scale of 70km. This distance gives a first indication of the scale at which clustering is forming.

With Ripley's K function confirming the clustering, measures to express the size of the clusters are considered by analysing the nearest distance to a neighbour, indicator variography, Gabriel graphs and Nearest Neighbour Graphs.

3.3.2 Nearest neighbour analysis

The first analysis was to subject the data to a nearest neighbour calculation as:

$$h = \left(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}\ \right)_{min} \ \forall\ \ i, j = 1, 2, 3, ..., n \ \text{ and } \ i \neq j$$

and logging the minimum distance h between two adjacent kimberlites.

Each of the datasets were analysed in turn, firstly all the data in Southern Africa (figure 3.4), then K1 (figure 3.5) and then K2 (figure 3.6).

The results are presented as a histogram of distances plotting the frequency of nearest neighbour distance on the primary axes, with the secondary axes (right hand side and top axes) showing a cumulative plot of the percentage of kimberlites falling within the search distance.

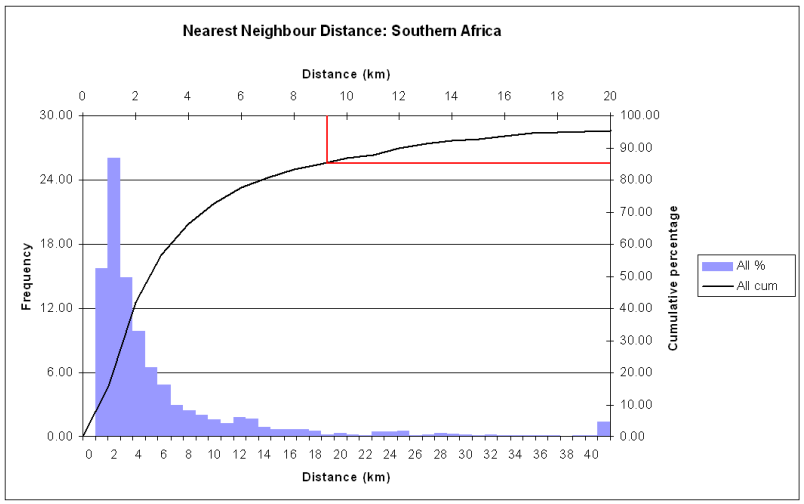


Figure 3.4: Nearest neighbour analysis: Southern Africa

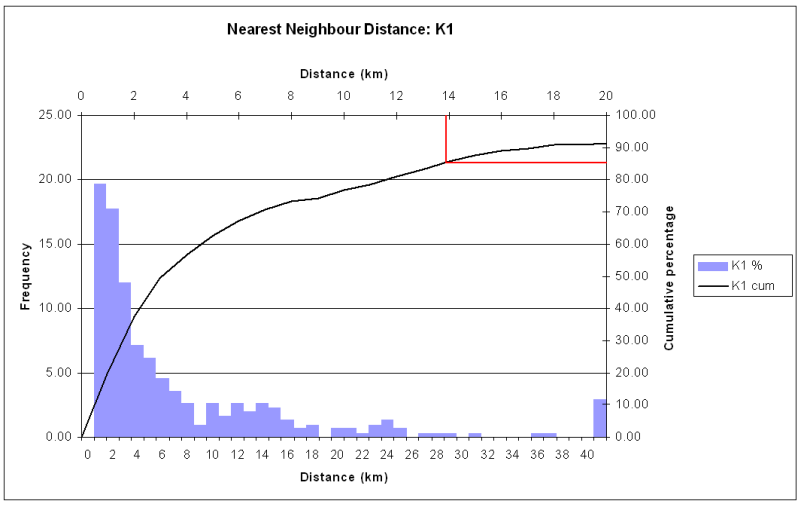


Figure 3.5: Nearest neighbour analysis: K1 kimberlites

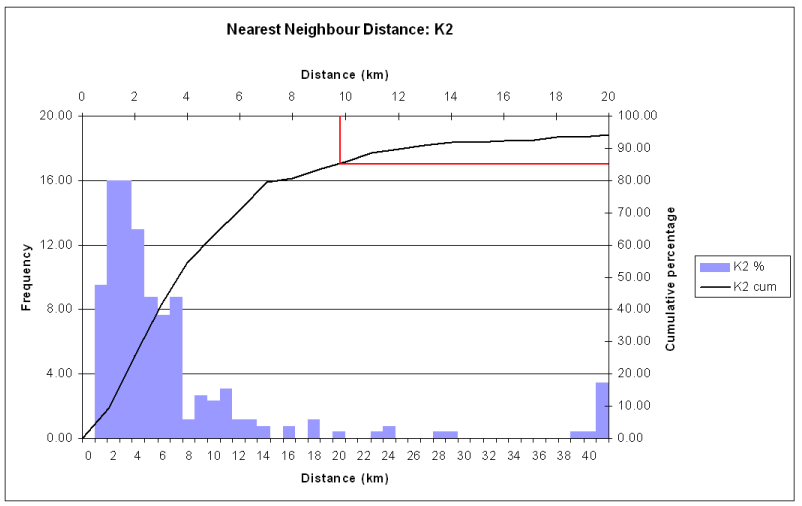


Figure 3.6: Nearest neighbour analysis: K2 kimberlites

Results

An examination of the results plotted on the secondary Y axis relative to distance and reading from the graphs shows around the modal interval that:

a) For a mixture of kimberlite types: 50% of the kimberlites are less than 2.4km apart
b) For K1 type kimberlites: 50% of the kimberlites are 3.1km apart and
c) For K2 type kimberlites : 50% of the kimberlites are 3.6km apart.

Depending on the level of confidence required, an appropriate distance can be determined from the graphs to deduce the probability of inclusion of kimberlites in a cluster.

Assuming an 85% probability for inclusion as acceptable, the distances are as follows (shown in red in figures 3.4 to 3.6).:

- A mixture of kimberlite types show a radius of 9.2km to be used
- If it is known that the type of kimberlite is K1, a radius of 13.8km should be used and
- If it is known that the type of kimberlite is K2, a radius of 9.8km should be used.

Studying the shape of the histograms for the craton area, K1 and K2, it appears as if a slight bimodality is observed at a distance of 10 to 14km. This suggests an interdependence between single kimberlites and adjacent clusters of kimberlites as a function of the spatiality measure.

In summary, usually at an early stage of exploration and target identification, the type of kimberlite is not known or the exact location of the pipe may not have been found in the cluster. In such a case, the longer of the distances (13.8km) should be used to secure an area around the midpoint of the target.

3.3.3 Indicator variogram analysis

With the kimberlite locations considered as a spatial variable, it is possible that an indication of the interdependence between kimberlite occurrences can be obtained using the range determined from an indicator variogram model.

The following data manipulation is proposed: Consider the kimberlite (X,Y) locations of all the kimberlites in Southern Africa. The immediate surroundings for each target area is carefully explored for the presence of other kimberlite occurrences and once a kimberlite is further developed, a series of holes are drilled around the kimberlite to ensure that infrastructure (recovery plants, town development, roads, etc.) are not accidentally placed on an adjacent orebody (condemnation drilling). Broadly, it can be assumed that within a certain distance around known kimberlites, no other undiscovered kimberlites will be found.

Assuming this distance to be 5km, based on the 50% coverage determined with the nearest neighbourhood calculations, and using the (X, Y) coordinates of each of the kimberlite occurrences, the following data representation is proposed:

Taking cognisance of the 50% inclusion of kimberlites:

On a grid of 1x1km blocks, for a surrounding grid distance up to 5km away, populate the grid

as follows:

$$\forall \text{ 1x1km location}(x_i, y_i) = \begin{cases} 1 & \text{if a kimberlite is present} \\ 0 & \text{with absence of kimberlite} \end{cases}$$

Graphically:

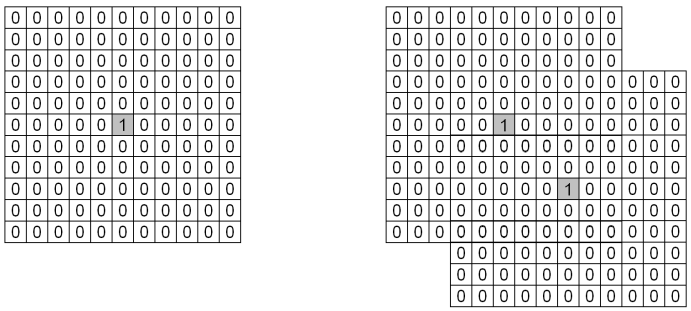


Figure 3.7: Populating 1x1km grid locations around three known kimberlite locations

Kimberlites very close to one another will have overlapping neighbourhoods. In such a case the maximum indicator is honoured with a value of one and the rest of the neighbourhood's bocks are coded with zeros. All other 1x1km blocks further than 5km from known kimberlites are coded as "missing" data. Figure 3.7(right) shows the 0 and 1 indicator blocks for two known kimberlites with overlapping neighbourhoods.

As a test to verify the stability of the variogram calculations, two regularised grids were created from this dataset. The grid containing $0/1$'s was regularised into blocks of 5x5km and 10x10km to represent the average number of kimberlites per 5x5km and 10x10km areas. Partially informed blocks with less than 10 observations were eliminated from the regularised dataset. The regularisation is used as a test to smooth the outcome of the variogram results from potential spurious effects. The exact modelling of all the parameters of the variogram is not required, as it is only the range parameter that will be used as a measure of spatial correlation.

Omni-directional variogram results are:

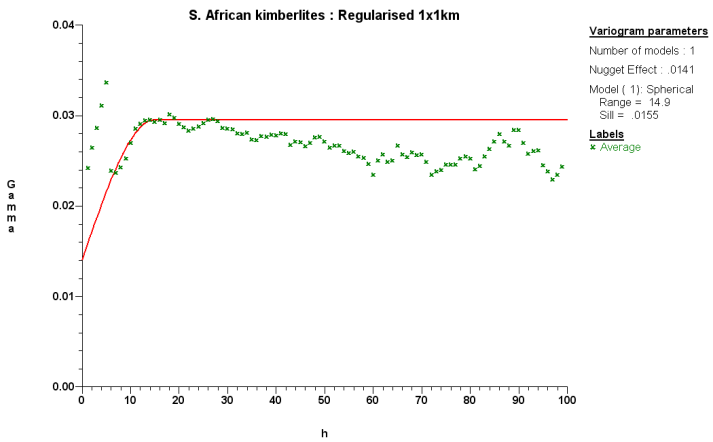


Figure 3.8: All Southern African kimberlites: 1x1km data

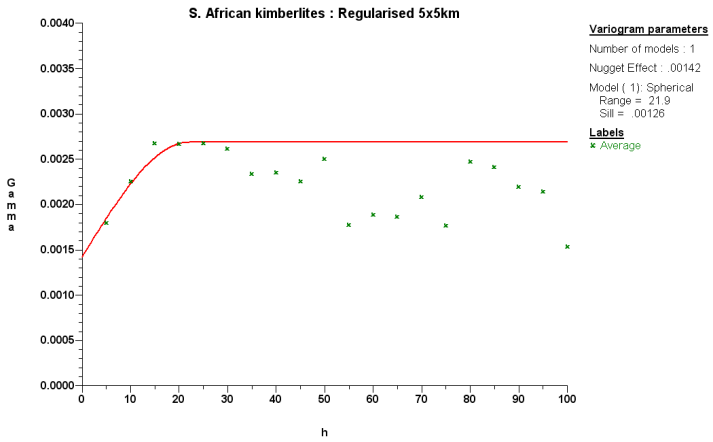


Figure 3.9: All Southern African kimberlites: 5x5km regularised data

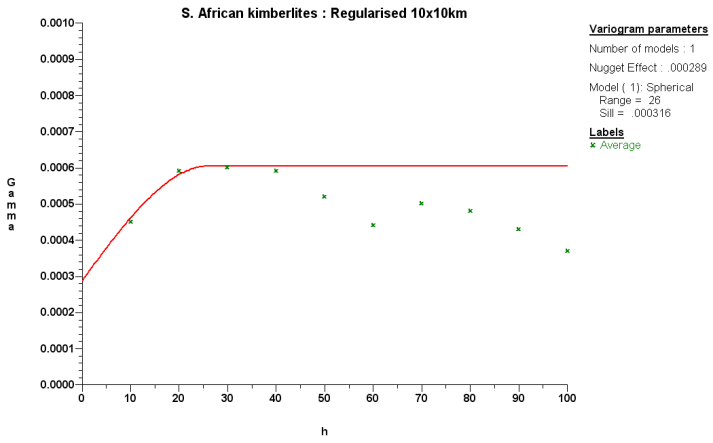


Figure 3.10: All Southern African kimberlites: 10x10km regularised data

Inspection of the results of the 1x1km data in figure 3.8 shows a linear feature for the first 5 experimental semi-variogram points. This is an artefact of the configuration of the 11x11 block of data surrounding each kimberlite and the variogram calculation's $1/(2\ n(h))$ factor.

The first five lags have a majority of zero values with a single indicator $I_{x_i,y_i} = 1$. As the distance between points increases and the number of pairs within the 11x11 block reduces, the factor $1/(2\ n(h))$ increases in $\gamma(h) = \frac{1}{2n(h)} \sum (Z_x - Z_{x+h})^2$. When the distance h exceeds the limit of a single 11x11 data configuration, or more than one kimberlite appear in the same configuration, normal variogram behaviour resumes.

This artefact disappears when the data is regularised.

Results

The ranges obtained from the modelled variograms are:

Indicator variogram range per dataset

Dataset	Variogram Range 1x1km data	Variogram Range 5x5km data	Variogram Range 10x10km data
K1	27km	25km	No fit
K2	32km	29km	No fit
S.Afr.	No fit	22km	26km

Although the clusters of kimberlites display prominent linear features on the craton, no preferential anisotropy on a local scale was observed in the variography.

From the data shown in the table, the variogram ranges would be expected to increase when the data is regularised, however, this is not apparent from the modelled results and the fitted ranges are all between 22km and 32km.

By eliminating the longest and shortest range values to eliminate the most extreme cases, the remaining data has an average range of 27km. This suggests that kimberlite locations are spatially correlated over a distance of 27km in Southern Africa, a measure that can be used in the determination of the size of an exploration target area.

3.3.4 Kimberlite clustering through proximity criteria

Various techniques exist that can be used to express a measure of clustering. Two methods for expressing the level of clustering of kimberlites were investigated:

a) The Gabriel graph (after Gabriel [67]) and

b) NNG, the Nearest Neighbour Graph [68].

By applying these methods to express the proximity or nearness of points, the (X, Y) locations of kimberlites are connected through a series of lines, based on rules formed by intersecting circles.

When the distances between the kimberlites are represented as a series of links sorted from the longest to the shortest and systematically eliminated, this expresses the breakup and clustering behaviour of the kimberlite locations.

Methodology

The dataset in Figure 3.11 below is used to illustrate the methods.

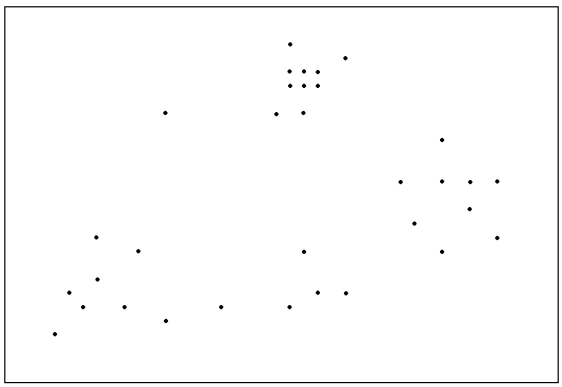


Figure 3.11: An example of kimberlite (X, Y) locations

A: The Gabriel Graph

The Gabriel graph considers the vertex set S with points P and Q within the set S, such that these points are adjacent if the line segment PQ represents the diameter of a circle containing no other elements of S. See Figure 3.12.

A single circle is constructed with midpoint $(x_{mid}, y_{mid}) = \left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}\right)$ and radius $r = \sqrt{(x_1 - x_{mid})^2 + (y_1 - y_{mid})^2}$. The circle is tested against all elements of S to ascertain if the segment PQ can be considered. If it is the case, the two points P and Q are connected.

Graphically:

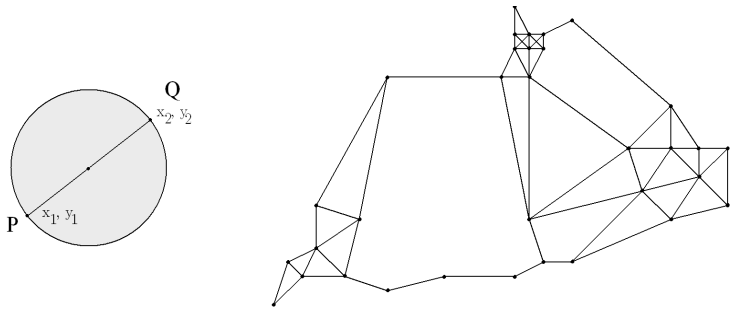


Figure 3.12: Gabriel graph - Linking nearest kimberlites

The set S of kimberlite connections are constructed for further analysis.

B: The nearest neighbour graph (NNG)

The NNG graph is the vertex set S with members of two overlapping circles constructed at P and Q. The circles are constructed with the first circle at midpoint P with radius PQ and likewise the second circle is constructed at Q with radius PQ. The vertex PQ becomes a member of S if no other members is found in the intersection of the two circles. See figure 3.13.

Schematically:

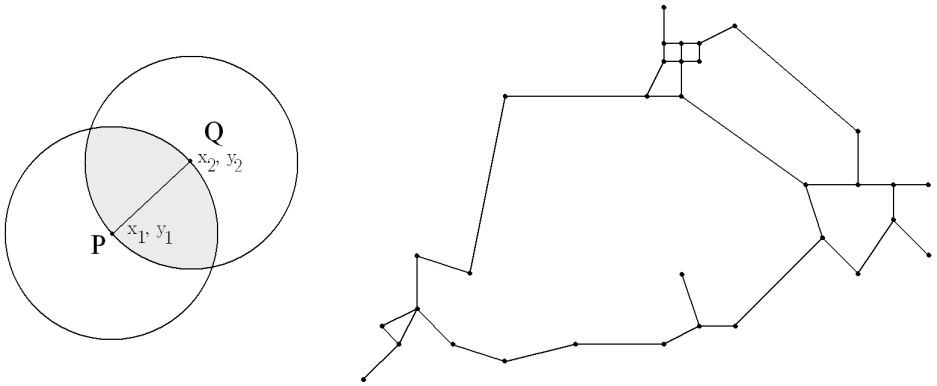


Figure 3.13: NNG - Linking nearest kimberlites

The set S of kimberlite connections are constructed for further analysis.

Determining cluster behaviour by systematic elimination of vertices from S

Once the vertices have been constructed, they are sorted from the longest to the shortest and systematically eliminated from the set. As the process proceeds, the network of connectivity will break up into clusters and single occurrences (singletons) of kimberlites. Statistics on the clusters and singletons are calculated as each distance is eliminated and cluster sizes and number of singletons are plotted.

Graphically:

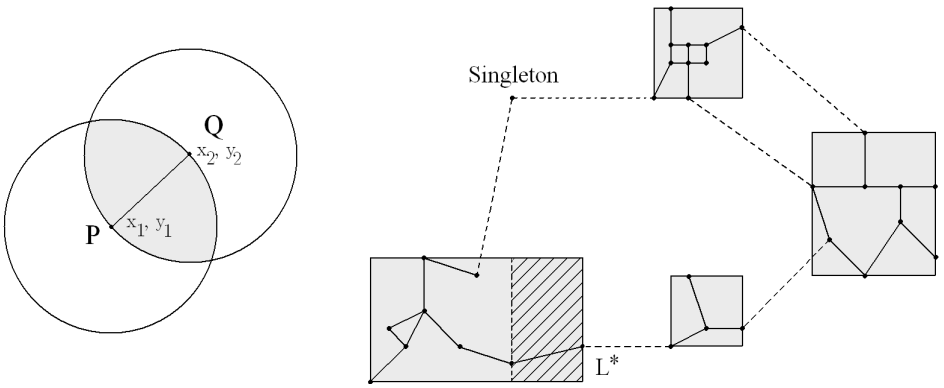


Figure 3.14: NNG: Eliminating links and calculating statistics on "clusters" and singletons. Link L^* and the hatched area will be eliminated next.

For each cluster of kimberlites, the minimum rectangular area that include all the points is quoted. A rectangular shape is used, as it represents a typical flight path pattern for an airborne survey. The elimination of links are continued until all points are presented as single occurrences and an analysis is done on the results to establish cluster behaviour. This will be demonstrated in the case study.

3.3.4.1 Case study

The Southern Africa, K1 and K2 datasets were used and analysed. The set S is calculated using the Gabriel and NNG graphs for each kimberlite type and is graphically represented in Figure 3.15.

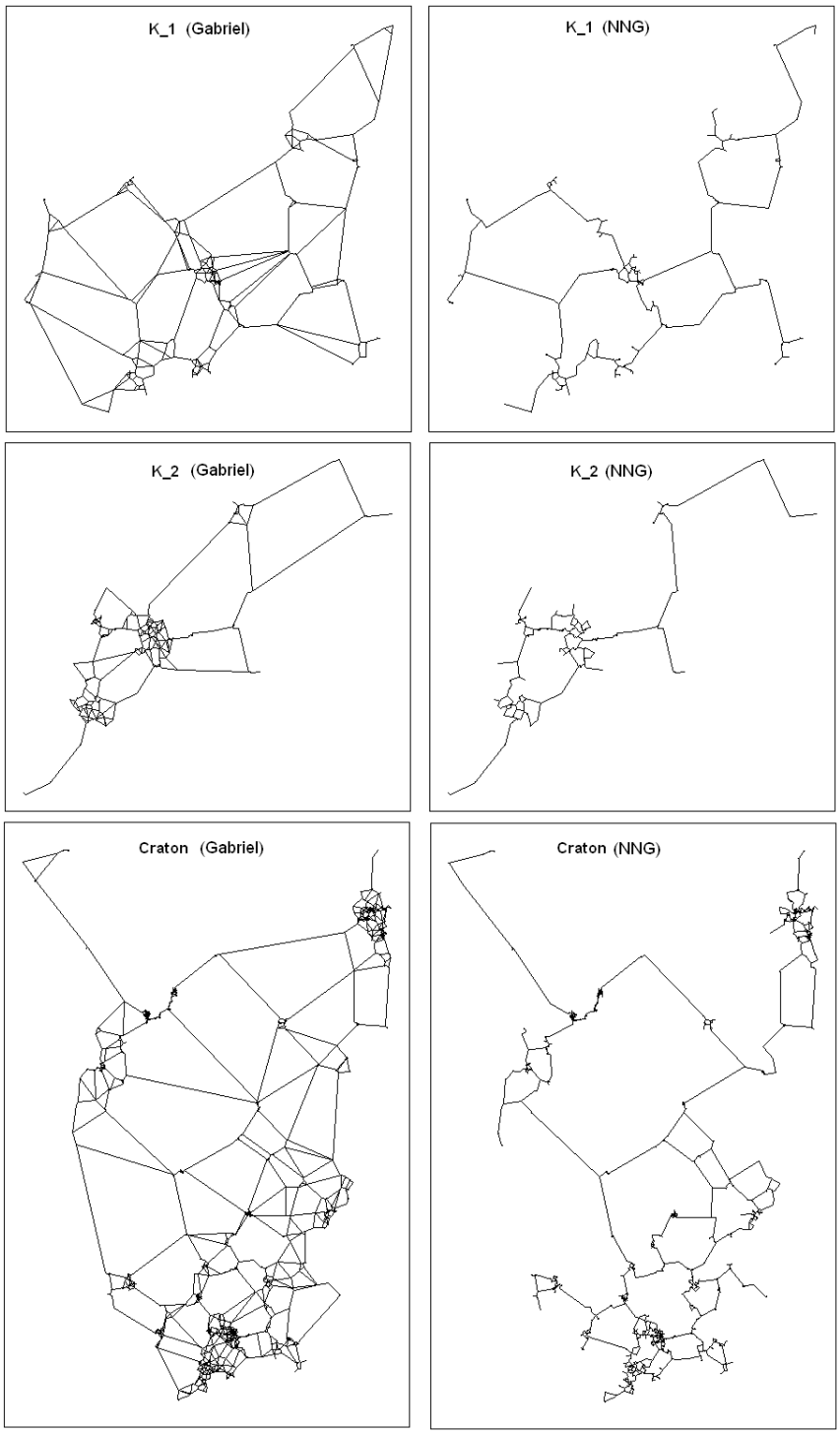


Figure 3.15: Connected lines for K1, K2 and all kimberlites of Southern Africa

Analysis of set S for Southern Africa kimberlites

A set S was determined for each of the networks and each contains distances sorted from the longest to the shortest, with a systematic elimination of the vertices.

A cluster is defined as a group of kimberlites containing two or more members with a positive area. As each link is eliminated, the following information were recorded:

- The distance eliminated (in kilometres)
- The number of clusters
- The individual and average size of all the cluster areas formed
- The number of singleton kimberlites

Histograms of this information are shown in figures 3.16 to 3.18.

a) The X axis is the length of the link eliminated (in kilometres).
b) Average area of the clusters formed after eliminating a distance
c) Frequency plot of the numbers of clusters at the eliminated distance
d) A graph of the frequency of singletons formed at the distance eliminated.

An interpretation is done to identify the distance, using the rate of increase in singletons as a guide, to identify the scale (length of link) at which clusters are breaking up. This implies that for a certain threshold of distance removed, the clusters that were "kept together" are now displaying "break-up" with the formation of smaller, and more clusters.

Sharp changes in the slope of graphs imply a sudden break-up of equidistant kimberlite occurrences. Points of interest are indicated with block arrows in the following graphs.

Results are as follows:

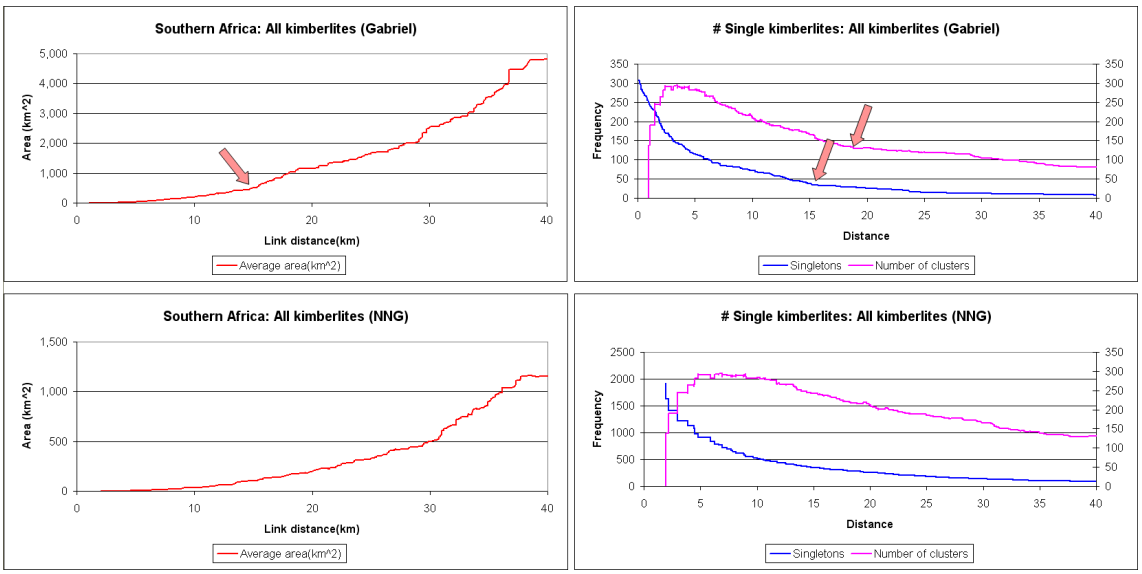


Figure 3.16: Results for SA kimberlites showing average cluster size (left) and rate of change in number of singletons and number of clusters versus distance (right).

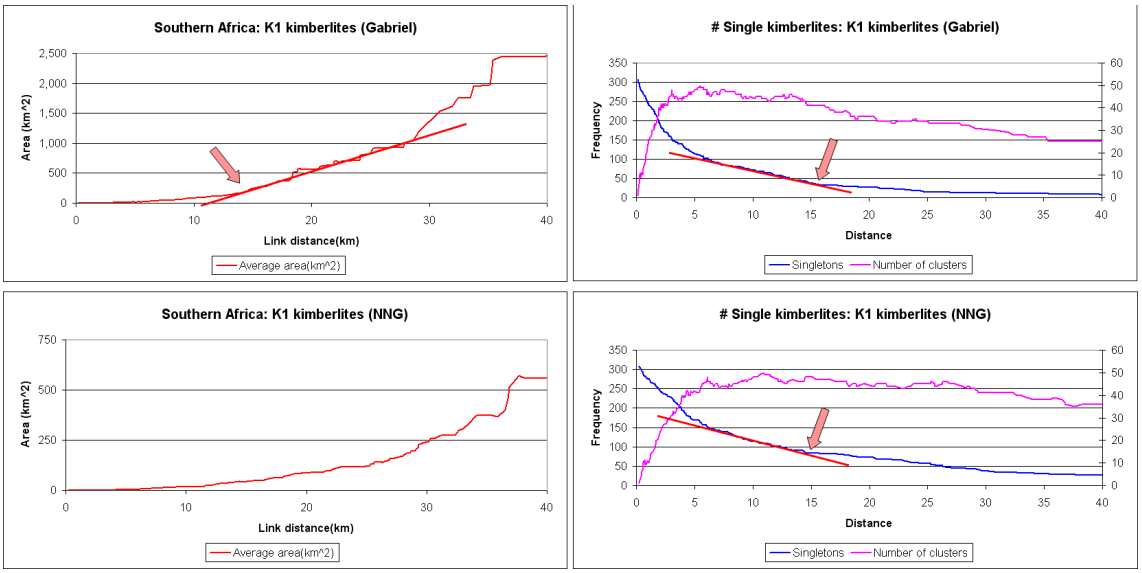


Figure 3.17: Results for K1 kimberlites showing average cluster size (left) and rate of change in number of singletons and number of clusters versus distance (right).

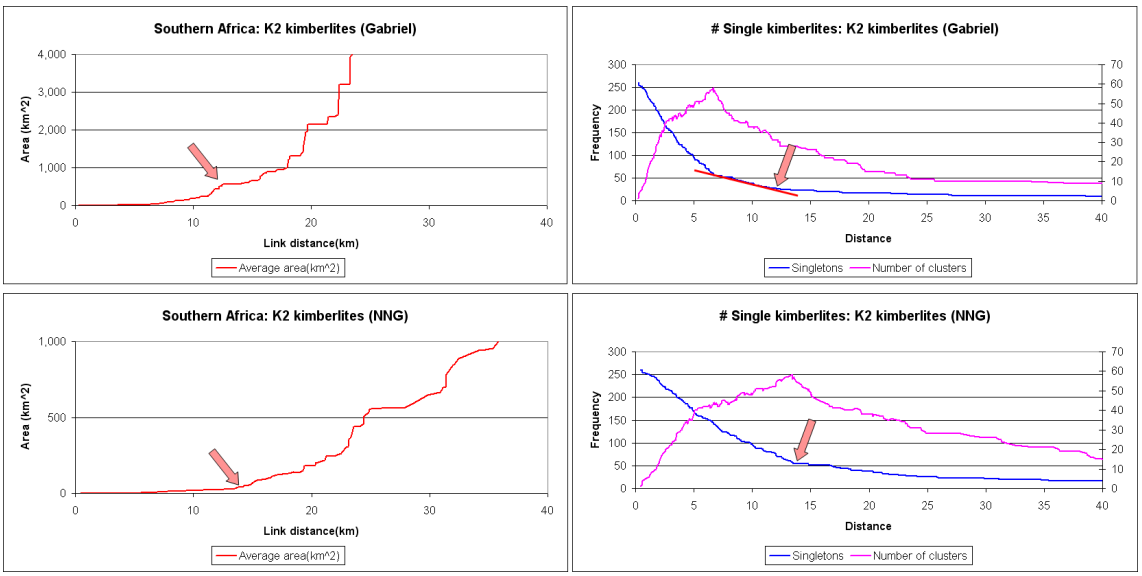


Figure 3.18: Results for K2 kimberlites showing average cluster size (left) and rate of change in number of singletons and number of clusters versus distance (right).

Examination of the figures and visually modelling straight lines through the singleton points, the following is observed.

From the rate of change in the plots of singleton kimberlite (as links are removed), for the K1 kimberlites, single kimberlites appear to form more rapidly from a distance of 15km, whereas

for K2 kimberlites, the distance is around 13km. Examining the rate of change in the singleton slope of the K1 graph relative to the K2 graph, shows that K1 is more dispersed and separates kimberlites more uniformly than K2. The slope of the curves suggests that the K2 pipes are more tightly clustered and further "break up" at distances around 7km. This is to be expected relating the geological behaviour of predominantly pipes (K1) relative to predominantly dykes (K2) kimberlite occurrences (personal communication with Dr. Field, Director/Consulting Geologist at DiaKim Consulting Limited).

The determined distances at which the singletons increase to indicate breakup of clusters are listed in the table:

Distance at which singletons increase

Method	K1	K2	All
Gabriel graph	$\pm$16km	$\pm$13km	$\pm$16km
NNG	$\pm$15km	$\pm$13km	unspecified

Examining the results for K1 & K2 shows the following:

K1: The distances where clusters break up are in the order of 15km

K2: The distances where clusters break up are 13km

Taking the average of the distances from the table above shows a breakup distance of 15km. Using this distance and calculating the average size of clusters shows that K1 clusters have an average area of 131km^2 and K2 clusters have an average of 260km^2.

During exploration, it is not always possible to know what type of kimberlite is in the proximity of the target area, thus the conservative approach would be to use the parameters obtained from the analysis of all the kimberlites (SA data) on the craton area.

The analysis and results indicate that by eliminating links at 15km, target areas of 182km^2 are formed which equates to a square area of: 13.5km x 13.5km[1].

Analysing $length/width$ ratio's of clusters

Assume a distance for cluster break-up of 15km, with results containing all the cluster configurations of the set S, with each of the members belonging to a cluster if single kimberlite occurrences are excluded.

A study of the average shape of the clusters was undertaken to ascertain whether the shapes are square or rectangular. Although the term cluster implies a circular or ellipsoidal shape, only square shapes are considered to ensure compatibility with the airborne geophysics flight paths. Histograms of the $log(width/length)$ ratios are shown:

[1]The target area is chosen to cover 85% of the kimberlites in the cluster to enable a comparison with paragraph 3.3.2

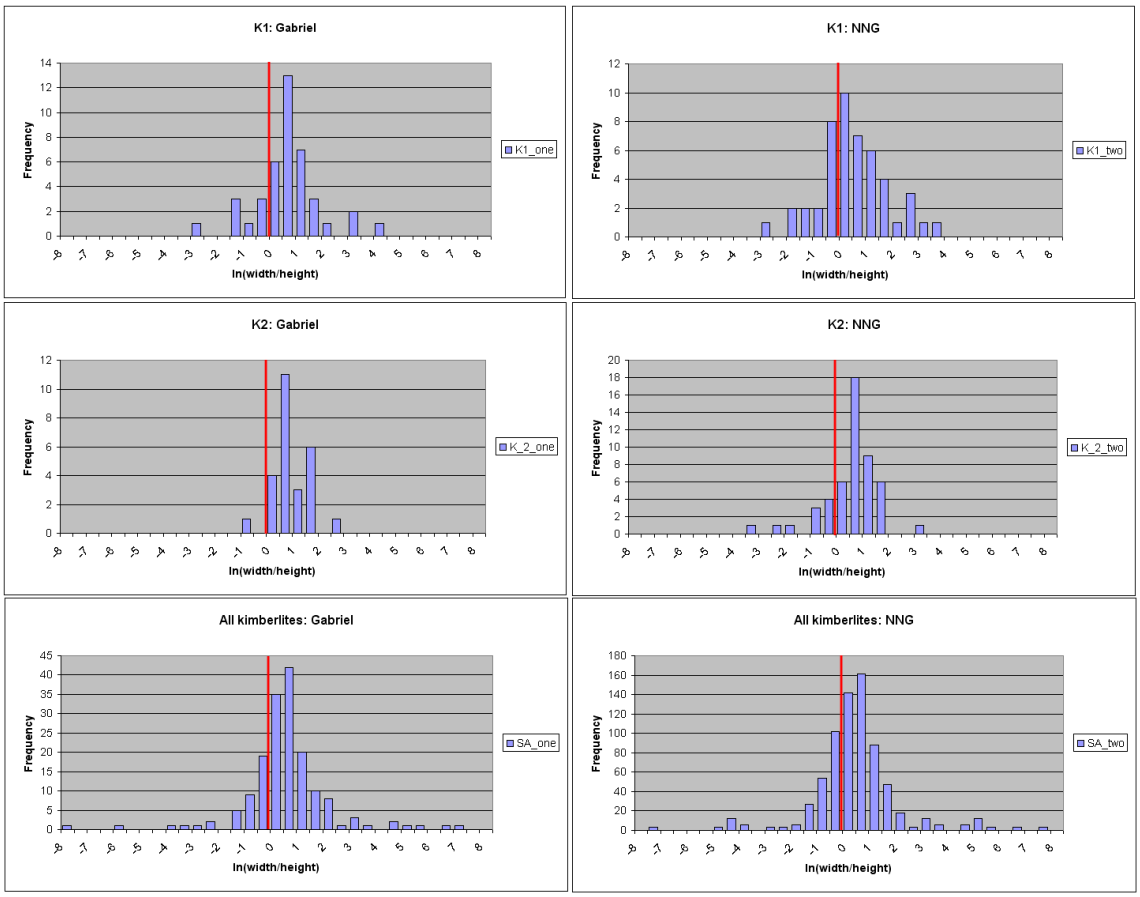


Figure 3.19: Anisotropy ratios expressed as $ln(width/height)$ for K1, K2 & All data

The mode of the histograms in figure 3.19 show a modal interval just above 0. This suggests that, on average, clusters are more elongated in the east-west direction than in the north-south direction. In practice, individual target areas will be dimensioned according to local geological features.

3.3.5 Analysis of results and comment

Distances calculated using the nearest neighbour method shows that 50% of kimberlites are within 3.1 to 3.6 km from one another and if an area containing 85% of a cluster's kimberlites is targeted, a distance of $\pm$14km is realistic to consider.

Although not directly related to cluster sizes, the spatial analysis shows that kimberlites are spatially correlated over a distance of approximately 27km.

Both the Gabriel and Nearest Neighbour Graph (as a subset of the Gabriel graph) have enabled the investigation of clusters at different scales. The building of the methods enables the link length and elements of the outcome to be manipulated and plotted as a combination of clusters and single kimberlites. This makes the interpretation of the behaviour of the kimberlite clusters possible.

The results of the two methods show that an area of 13×13km can be considered to cover 85% of the kimberlites in a cluster. The geology will differ from one area to another, but lacking local interpretation of geology may favour a slight bias to a target area more extended in the east-west direction than the north-south direction.

Figure 3.20 shows the location of kimberlite clusters after links with lengths greater than 15km have been eliminated.

These results are only applicable to the Southern African continent and should not be applied to other areas. Southern Africa remains an area with extensive exploration potential and the results of this study have provided an indication of approximate size of target areas.

3.4 Analysing indicator mineral data to identify kimberlite targets

3.4.1 Data

Indicator mineral data obtained from two projects (De Beers Canada Inc: Exploration and De Beers Exploration: South Africa) were used for this study.

The data covers large areas and at each sample location, measurements were made on the soil based on visual and/or chemical analyses of the indicator minerals. The soil samples were also geologically categorised to describe the various environments in which the indicator minerals were found.

The Canadian dataset comprises 97 variables per sample location at ±36 700 sample localities and the South African dataset comprises ±129 000 samples, each with 15 variables.

The response variable will be formulated as a categorical variable based on the data at known locations, as a distance class to the nearest kimberlite.

3.4.2 Objective

The objective of this research is to determine if it is possible to identify new target areas through (geo)statistical analysis of the indicator mineral results.

With the locations of the soil samples known the possibility exists to do a spatial as well as a statistical analysis.

New kimberlite anomalies will not be found using only the geostatistical interpretation nor with only using statistical predictors. The aim of this research is to identify a method that could combine the spatial character of the data with a statistical multivariate predictor model.

Several data mining methods were investigated to find the best model. The classification tree method, as described by Breiman et al. [10] showed the most promising results and was accepted as the method of choice. The output of the model is easy and practical to interpret and was applied to two case studies.

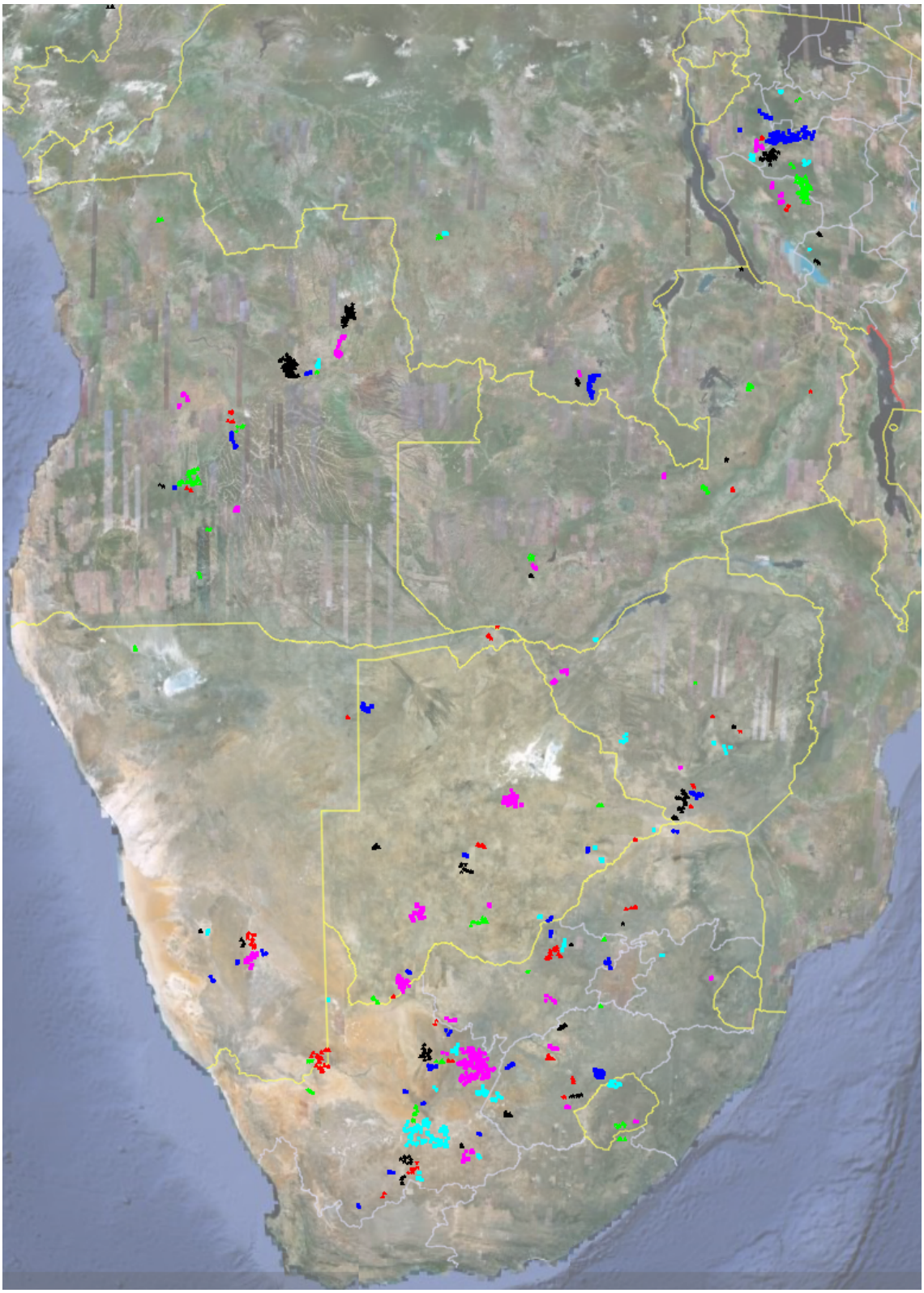


Figure 3.20: Cluster locations (eliminating links greater than 15km) with different clusters represented by different colours

3.4.3 Classification trees: A datamining method

Classification trees is a method used to predict membership of cases into classes of a categorical dependent variable from measurements of one or more predictor variables [2].

The purpose of using and building a classification tree is to be able to use the measurements from soil samples to model their proximity to a kimberlite pipe.

Relating the definition of a classification tree to an application for kimberlite exploration, is to use multivariate measurements (as predictor variables) taken from soil samples collected over many year from a large exploration target area, to model the distance of each sample (or case) relative to known kimberlites. The distance between a sample and the nearest kimberlite is calculated and categorised into a class which is used as the categorical dependent variable. An example of class intervals is: 0-5km, 5-10km and >20km.

The predictor variables consists of frequency counts of the different chemical grains present in the soil samples associated with kimberlitic material. A higher frequency count indicates a closer proximity to a kimberlite pipe. The counting of grains is stopped once a certain number of grains had been counted and what makes the data challenging to use is that this threshold is inconsistent from one sampling campaign to the other.

Once the model is built for kimberlites at known locations, it is applied to all the data to find kimberlites at unknown locations by classifying each sample with a distance class. The location of samples classed in close proximity to a kimberlite indicates a potential target area that can be explored in more detail. This is demonstrated in the case studies.

Once the model is built, it is applied to all the data to associate each sample with a distance class. The locations of samples classed in close proximity to a kimberlite becomes a target area for follow up exploration work. This is demonstrated in the case studies.

3.4.3.1 Model building and methodology

The model building starts off with all the data in a parent node (the root node) which is split into two data groups as child nodes through the use of a chosen variable and threshold value. A misclassification rate is used as a measure to perform this split to determine which variable and at what threshold value the misclassification of cases is at a minimum.

For example, the kimberlite case study with three distances (a three class problem with classes 1, 2 and 3) is considered. Suppose that after a two-way split of the root node cases there are n_1 cases in the first child node and n_2 cases in the second child node. The first node has each of the three distance classes present in proportions p_1, p_2 and p_3 and the second child node has the classes in proportions q_1, q_2, and q_3. For the purpose of classification of future points (of which only the predictors are known, but not the distance from the nearest kimberlite), each node is assigned the class that yields the largest class proportion from the training data. If, for the first child node, class 3 is in abundance with relative frequency p_3 ($p_3 > p_1$ and $p_3 > p_2$), the node will be used to identify class 3. Future cases that belong to classes 1 or 2 but land up in the first child node when the tree is applied, will be misclassified. Similarly, the second child node will have a class associated with it, say class 2, with the cases in proportions

[2]Internet, http://www.statsoft.com/textbook/classification-trees, accessed: July 2010

q_1 and q_3 (or 1-q_2) misclassified.

The Overall Misclassification Rate (OMR) for the two child nodes at the split under discussion in the three class problem is defined by weighing the misclassification probabilities proportional to the number of cases in each node:

$$OMR = \frac{[n_1(1-p_3) + n_2(1-q_2)]}{(n_1+n_2)}$$

Brute force calculation by computer iterates through all variables and all splitting values for each variable to find the predictor variable and its best splitting value that minimises the OMR. This is done as follows: The algorithm takes one of the predictors, say X and selects a splitting value t. For each of the cases in the parent node with predictor value $X <= t$, the case is moved to the first child node left (say L), else to the second child node right (say R). This is done for all cases in the parent node. The OMR is then calculated. After considering every value for t and each predictor variable X, the predictor variable and its t value yielding the smallest OMR is used to produce the first binary split of the parent node. The same procedure is repeated for each of the child nodes and each time the binary split with the smallest OMR is used to split the data into further levels of child nodes.

Figure 3.21[3] graphically demonstrates the outcome of 3 class problem with binary splitting.

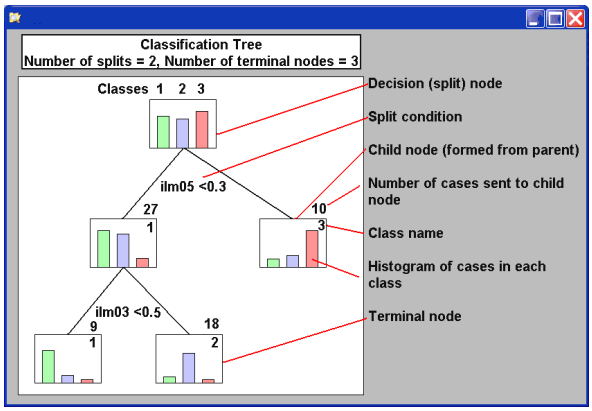


Figure 3.21: Example demonstrating a classification tree model building process

The process can be repeated until each terminal node contains a single case in which case the misclassification rate, defined as 1 minus the proportion of correctly classified cases ($MCR = 1 - p_i$), is zero. To continue to this point is meaningless and there is obviously no benefit in doing this as such a model is grossly over fitted and unusable. A stopping criterion is needed to trim this model building process.

This is done as follows: At any stage of growing a tree T with M terminal nodes, the overall tree impurity is calculated by weighing the MCR for terminal node m, $MCR_m(T)$ with the number of cases in the node, say N_m relative to the total number of cases N as follows:

$$MCR(T) = \sum_{m=1}^{M} N_m MCR_m(T)/N$$

[3]Based on figure from http://www.statsoft.com/textbook/classification-trees, accessed: July 2010

The right sized tree is determined by minimising a cost penalty criterion $C_\alpha(T)$ for a chosen tuning constant α (subjectively chosen with $0 < \alpha < 1$) and number of terminal nodes (M) present in the tree. This criterion is dependent on and penalised with increasing number of terminal nodes as:

$$C_\alpha(T) = MCR(T) + \alpha M$$

The $MCR(T)$ is a decreasing function of M and penalising it proportional to the increasing M, $C_\alpha(T)$ has a minimum value reached at the tree with M^* terminal nodes. This gives a first indication of the right-sized tree to use. Some subjectivity is used for the final selection of the model. By looking at the complexity of the tree structure and through inspection, comparing one tree to the other, the success rate of the model applied to the test data is used to choose the final model.

The outcome of the tree produces a model through which a new case can be filtered downwards until it falls in a node with an associated class value.

For the kimberlite case studies, this filtering is done on all the datapoints falling outside the immediate surrounding area of the known kimberlites. Classifying these points will categorise their expected distance from an unknown kimberlite into one of the three distance categories. If a cluster of points are found indicating a close proximity of a kimberlite, this would be a target for further exploratory investigation.

The process is numerically intensive and the software package Statistica was used to perform the analysis.

3.4.3.2 Interpreting Statistica output

Example output from Statistica shows how a progressively more complex tree structure has a reduction in cost with an associated increase in terminal nodes. See figures 3.22 and 3.23.

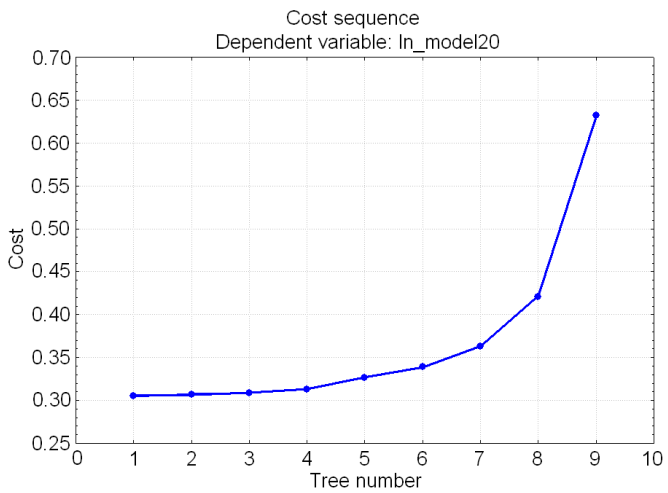


Figure 3.22: Cost sequence of trees.

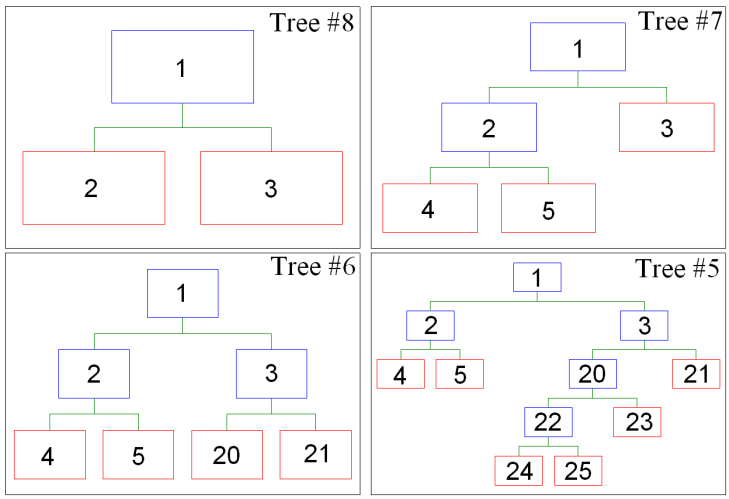


Figure 3.23: Tree graphs showing terminal nodes and splits (Tree #8, #7, #6, #5)

By inspection of the diminishing reduction of the cost function beyond tree #7 and the increase in the complexity of tree #6 and beyond, the choice would be to accept tree #7 as the model. Once accepted, Figure 3.24 below summarises the success rate of this model as applied to the test dataset. For this example, the green shaded blocks show a very low percentage of predictive success for cases with a distance class of $< 5km$ and such a model would need to be refined further.

	Classification matrix 2 (11Canadian data 2006 (ALL data).sta) Dependent variable: ln_model20 Options: Categorical response, **Test sample** Include condition: ln_model20 > 0				
	Observed	Predicted 5km	Predicted 10km	Predicted 20km	Row Total
Number	5km	5	21	171	197
Column Percentage		31.25%	36.84%	31.90%	
Row Percentage		2.54%	10.66%	86.80%	
Total Percentage		0.82%	3.45%	28.08%	32.35%
Number	10km	4	20	150	174
Column Percentage		25.00%	35.09%	27.99%	
Row Percentage		2.30%	11.49%	86.21%	
Total Percentage		0.66%	3.28%	24.63%	28.57%
Number	20km	7	16	215	238
Column Percentage		43.75%	28.07%	40.11%	
Row Percentage		2.94%	6.72%	90.34%	
Total Percentage		1.15%	2.63%	35.30%	39.08%
Count	All Groups	16	57	536	609
Total Percent		2.63%	9.36%	88.01%	

Figure 3.24: Demonstrating predictive power (green) and miss-classification (encircled).

The outcome of the modelling process is a tree structure that can be applied to the full dataset to highlight potential kimberlite target areas.

This is demonstrated in the following two case studies.

3.4.4 Case studies

Two case studies are presented for which a classification tree analysis was done, resulting in the identification of kimberlite anomalies.

The first study is on data from Canada and the second from South Africa.

In both studies, many data points were available, but not all fields were populated and this meant that a number of challenges had to be overcome in the data preparation stage.

3.4.4.1 Case Study 1: Analysis of Indicator Mineral data from Canada

A total of 36 757 data points were available for the Canadian study and Figure 3.25 is a plot where each black dot represents a sample location. The blue circles indicate known kimberlite locations and arrows show the directions of movement of glaciers that covered Canada many years ago.

The data file contains 97 variables, of which those variables that were populated and most relevant to the study were identified by the project geologist for analysis. Two general categories are assigned to the samples, namely visual data and geochemical data. A subset from the variables containing mostly visual data is shown in table 3.1, with summary univariate statistics:

Var. Name	n	Mean	Variance	Min	Max	Range	#missing
LONG X	36757					1537354	0
LAT Y	36757					1116956	0
VOLUME	36757	23.632	11748.167	3	9680	9677	0
H0	13348	0.306	19.231	0	101	101	23409
J1	13348	0.778	12.900	0	162	162	23409
J4	13348	1.827	56.703	0	132	132	23409
J5	13348	1.720	50.554	0	188	188	23409
K ILM	13348	3.941	241.099	0	339	339	23409
Gar-03	28297	0.187	3.658	0	105	105	8460
Ilm-03	28297	1.266	65.805	0	453	453	8460
Ilm-05	34285	1.518	7188.890	0	15526	15526	2472
OTH-05	10715	0.207	42.700	0	661	661	26042
Ilm-tot	36755	2.458	7653.089	0	16474	16474	2
TSP-tot	36755	1.204	42.662	0	382	382	2
Olivine-03	10722	1.513	321.241	0	1677	1677	26035
OPX-03	9915	1.640	410.097	0	1677	1677	26842
Olivine-05	9002	0.121	4.177	0	100	100	27755
Olivine-tot	10727	1.614	334.167	0	1677	1677	26030
OPX-tot	10120	1.973	406.730	0	1677	1677	26637

Table 3.1: Summary Statistics for visual variables

The data were taken from different geological domains, categorised into: Glaciofluvial, Till, Stream, Beach, Ablation Till, Basal Till, Glaciolacustrine and other geologies where the geology is unknown. Each sample is sieved into size categories and if indicator minerals are present they were chemically identified, counted and logged against the size fraction.

The data comprises visual counts in size fractions of 0.3mm and 0.5mm and counts of chemical types. Variables used in the analysis were measurements of Garnets (visual data and mineral chemistry: H0-H6), Ilmenite (visual and chemical: J1 to J6), kimberlitic total Ilmenite (J3 to

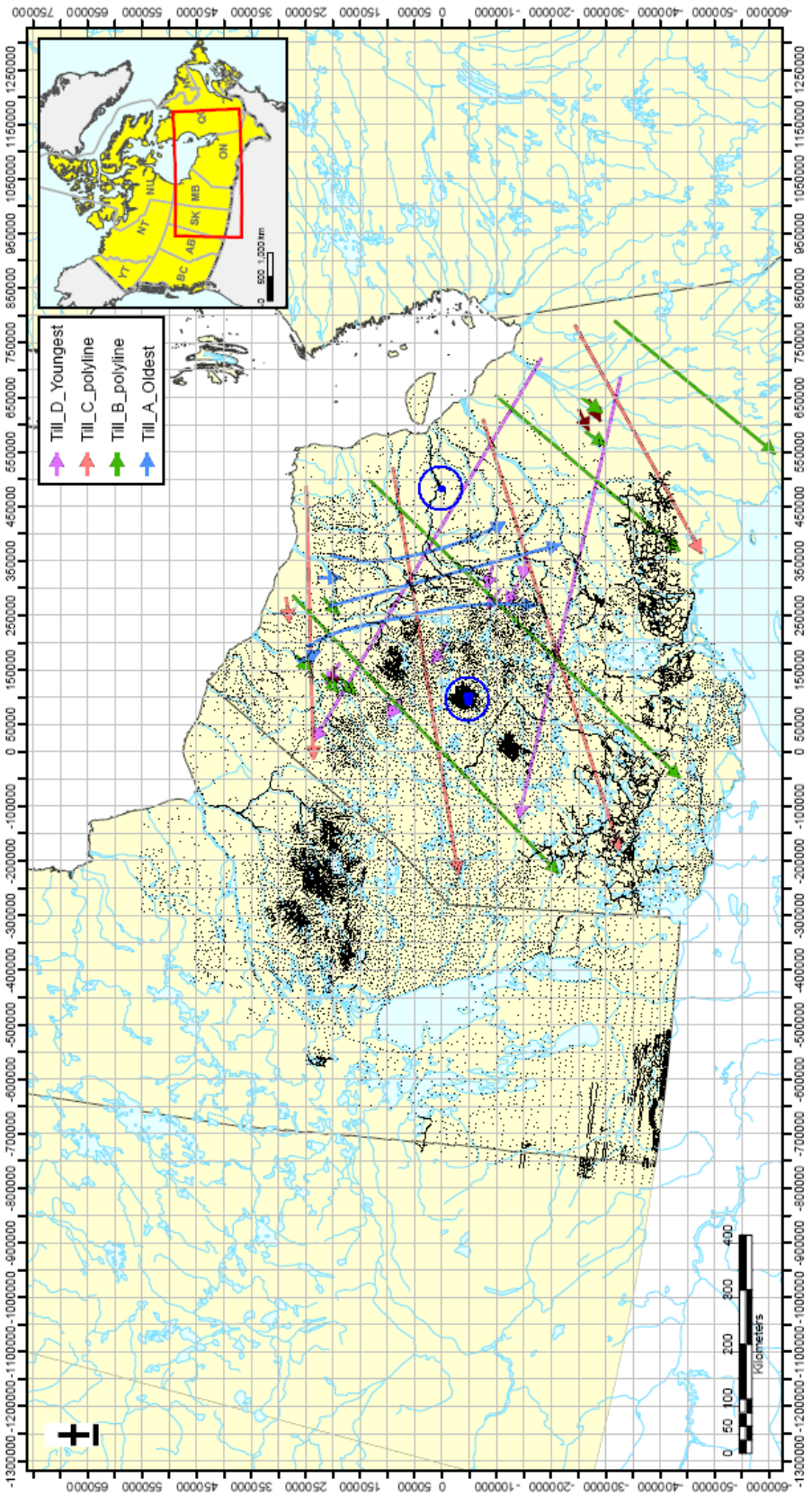


Figure 3.25: Location plot with Indicator Mineral samples in Canada

J6): K-ILM, Spinels, Ortho pyroxene: OPX, Clinopyroxene, Olivine and various combinations of totals for the variables.

There are large numbers of missing values in the data and in Figure 3.26 a plot was made (in black) for each variable containing a missing value. The X axis is the variable name and the Y axis represent a thin black line for each of the 36 757 samples that contain a missing value. The black area represents missing values and is significant. It is noteworthy that the missing values are not consistent across a sample case, as some variables may have a count for the 0.3mm size class, but may have a missing value for the 0.5mm size fraction.

Preparation for classification tree analysis

A significant amount of data preparation and manipulation is required on this data in order to carry out a classification tree analysis and a summary of the most pertinent data challenges are given:

Missing data: Missing data appear in large numbers scattered throughout the dataset. Lines of data contain missing values inconsistently from one variable to the other.

Sieving sizes changes: Sample campaigns in some years utilised sieve sizes of 0.3mm, 0.5mm and 1mm sizes, but in other years, the sizes were 0.3mm, 0.425mm, 0.5mm, 0.71mm or combinations of these

Weight measurements: Missing weight measurements resulted in challenges standardising the counts per unit volume, and assumptions about weight had to be made.

Perceived reliability of data: Data from the different years was classified into low, medium and high confidence categories but could not be quantified. These classifications were ignored during the model building process.

Truncations of counts: Counts of particles were capped during the data collection phase and truncated once a certain count was exceeded, but the truncation values were not always consistent and varied between sample programmes as either 30, 50 or 100 particles. The truncation was also different between variables. No measurement was made of the percentage weight of the sample sorted, and could therefore not be used to estimate the total count per sample.

Once the data had been prepared, the model building process was started.

Creation of class variables

Three distance classes were used for the application of classification trees in both case studies. Several iterations of distance classes were attempted to obtain sensible models and eventually two sets of class distances were fully analysed, with the predictor being a class variable describing the distance a sample is from a known kimberlite.

The two sets of class indicators created for the data surrounding known kimberlites were:

a. Set 1: Classes for 5, 10 and 20km from a kimberlite and
b. Set 2: Classes for 2, 5 and 10km from a kimberlite

Figure 3.26: Graphical representation of missing values (visual data)

For each sample, the distance from a known kimberlite pipe is coded by an indicator. For example, the indicators for the (5, 10 and 20km) set is:

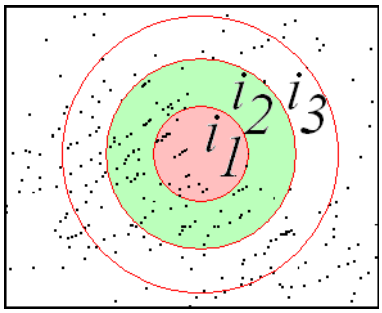


Figure 3.27: Definition of indicators

$i_1 = 1\ \forall$ samples $\leq$ 5km from kimberlite, else $i_1 = 0$
$i_2 = 1\ \forall$ samples $\leq$ 10km from kimberlite, else $i_2 = 0$
$i_3 = 1\ \forall$ samples $\leq$ 20km from kimberlite, else $i_3 = 0$
$i_4 =$ "unknown" (for all other data points)

The data was split into a training and a test dataset (respectively 1528 and 491 observations each) using only the data immediately surrounding the known kimberlites. The classification tree fitting is undertaken using the training data and the method builds a model to identify distance categories that is based on visual and geochemical data.

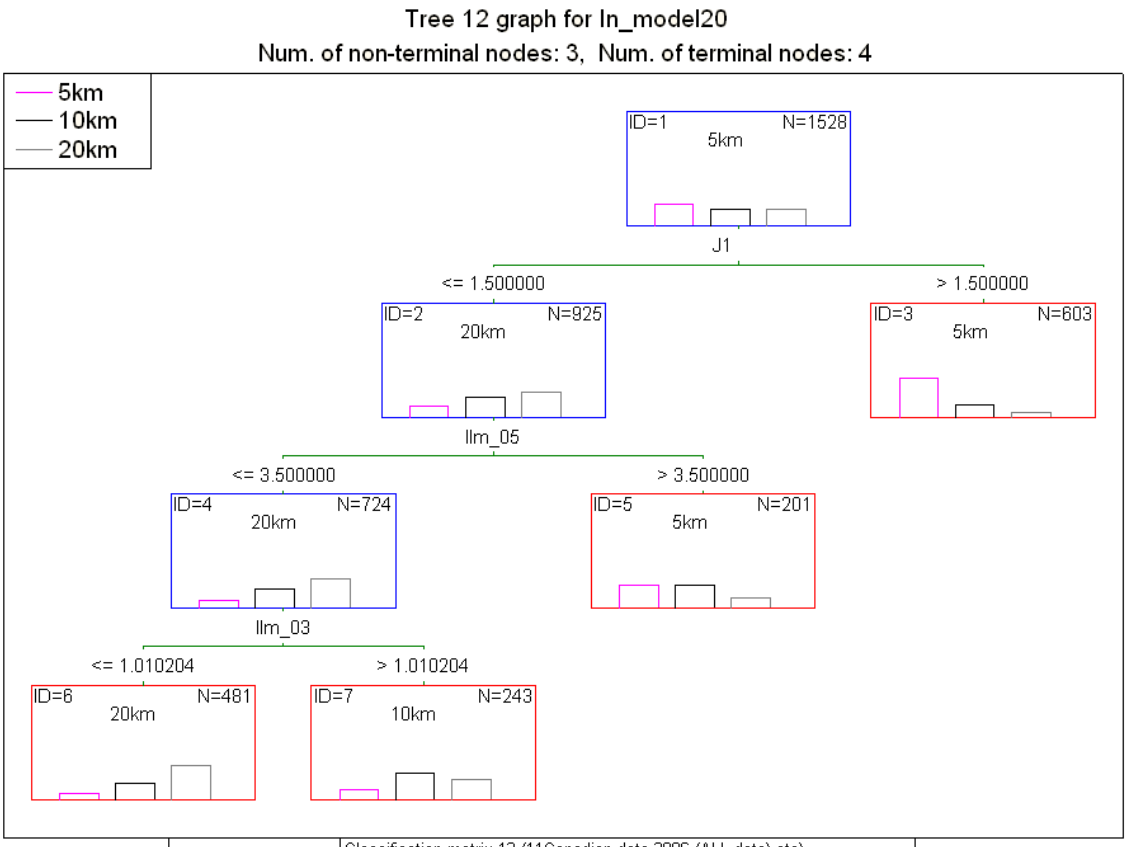
Once the model has been determined and tested for acceptability using the test dataset, the "$unknown$" cases were subjected to the model and categorised. The resultant distant categories obtained from the model were plotted in different colours representing the predicted distance from a potential kimberlite target. See Figure 3.29.

The results of the process were as follows:

Resultant classification tree models

Model for (2, 5, 10km): This model did not give satisfactory results based on the statistics obtained from the test dataset and could not be used to predict the proximity of kimberlite targets with sufficient confidence. This model was abandoned.

Model for (5, 10, 20km): The results for this model were satisfactory when tested on the test dataset and the tree with results are shown in Figure 3.28.



Classification matrix 12 (11Canadian data 2006 (ALL data).sta)
Dependent variable: ln_model20
Options: Categorical response, Test sample
Include condition: ln_model20 > 0

	Observed	Predicted 5km	Predicted 10km	Predicted 20km	Row Total
Number	5km	156	12	21	189
Column Percentage		57.35%	15.58%	14.79%	
Row Percentage		82.54%	6.35%	11.11%	
Total Percentage		31.77%	2.44%	4.28%	38.49%
Number	10km	76	33	29	138
Column Percentage		27.94%	42.86%	20.42%	
Row Percentage		55.07%	23.91%	21.01%	
Total Percentage		15.48%	6.72%	5.91%	28.11%
Number	20km	40	32	92	164
Column Percentage		14.71%	41.56%	64.79%	
Row Percentage		24.39%	19.51%	56.10%	
Total Percentage		8.15%	6.52%	18.74%	33.40%
Count	All Groups	272	77	142	491
Total Percent		55.40%	15.68%	28.92%	

Figure 3.28: Classification tree model for predicting sample proximity to kimberlite targets

The model was reasonable for predicting the proximity of a kimberlite to a sample at a distance of 5km (based on the test dataset: 82.5%) and was deemed adequate. For this distance, the predictor model is: Per litre of sample material, the J1 count must be more than 1.5 or, if J1 $<$ 1.5 but Ilm-05$>$3.5, a sample could be within 5km of a kimberlite target with 82.5% probability.

Results and IM targets predicted by applying the model to the unclassified data, using the classification tree from Figure 3.28, are shown in Figure 3.29:

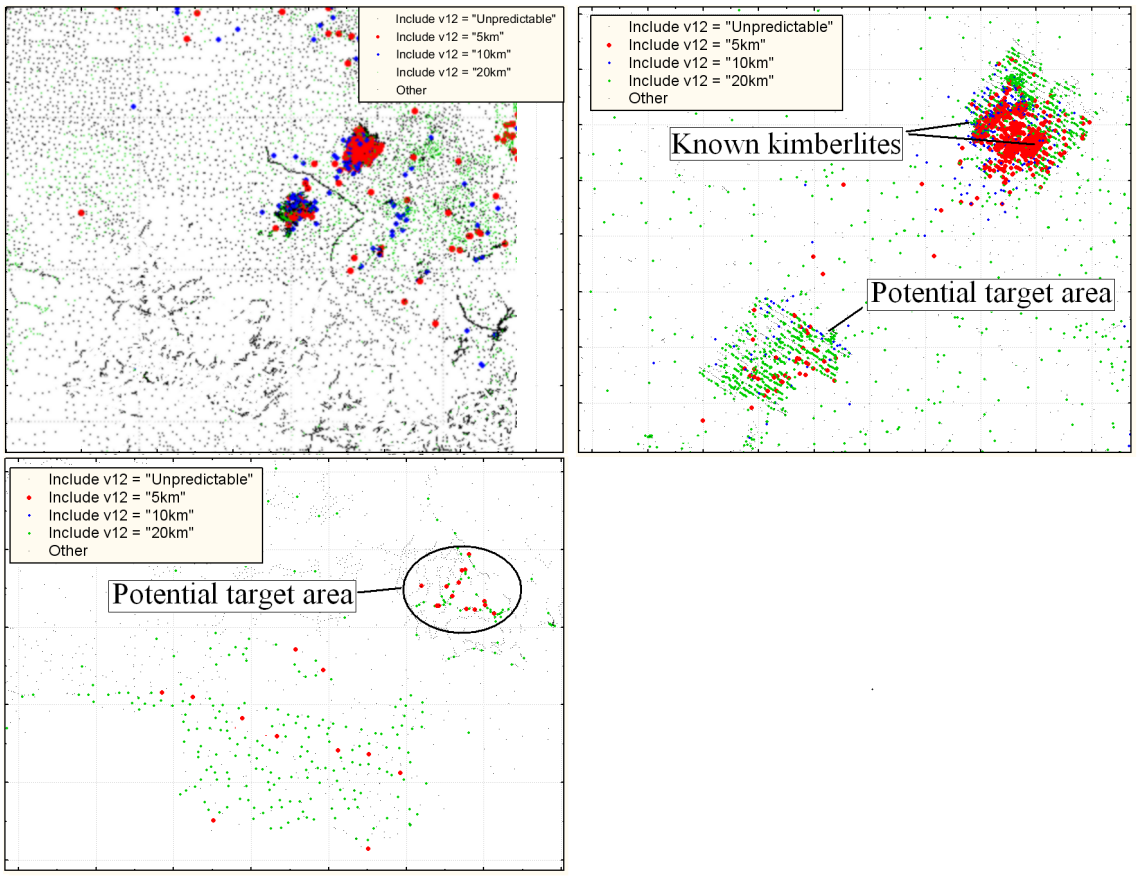


Figure 3.29: Results: Locations plotted in red indicate areas with potential kimberlitic anomalies

Results

The areas indicated by groupings of red indicators highlights potential kimberlite targets within a 5km distance. It is evident from Figure 3.29 (top right) that a densification of samples happened around one such a potential target.

However, care must be taken with the interpretation of the data and model because ice movement (multiple events from different global and local directions) would have transported the indicator minerals in specific directions and the direction of the last glacier would dominate the behaviour. In this analysis, a simplified model, not taking the ice movement into consideration was assumed for the model building in a first attempt at identifying targets and applying classification trees as a model. Ideally the ice flow should be taken into account where possible. Once the most relevant variables have been identified in a first model, a refinement of the model should be attempted by defining the indicator variables according to the ice directions. See Figure 3.30.

Graphically:

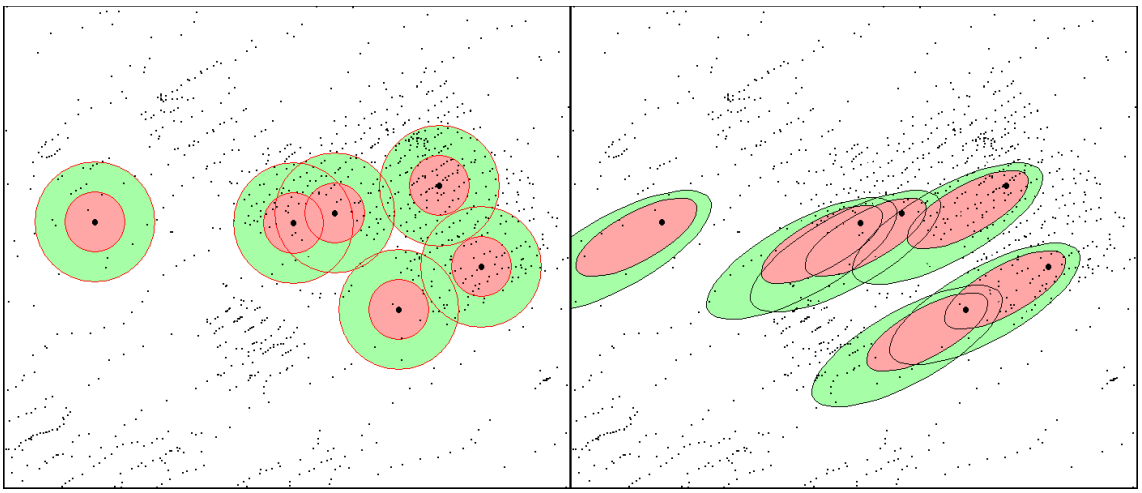


Figure 3.30: (Left): Simplified case showing the first two indicator classes (Right): More ideal classification incorporating local glacial movement

Formally, only regional ice direction movements are known and are not defined on a local scale.

3.4.4.2 Case Study 2: Indicator Mineral analysis for data in North West province : South Africa

The methodology used to evaluate the Canadian data was followed for the South African data but the application was not as successful.

The reason is that the kimberlite types that were used for the training data are derived from distinctly different types of kimberlites and the indicator mineral footprint does not display homogeneous characteristics. This makes the training and model building of the classification tree more difficult and results in a higher rate of misclassification. Even with the shortcomings of the model, new potential targets were highlighted for further consideration.

Summary results are presented for this case study and are shown in Figure 3.31:

Discussion and results

Interpreting results from Figure 3.31:

• The top left figure shows the sample locations (soil samples) from which the indicator minerals were recovered.
• Top right shows the results from the model considering the 2, 5 and 10km model
• Bottom left shows the results from the model considering the 5, 10 and 20km model. The test results from fitting a classification tree was not very successful, but still deemed acceptable for use in the identification of potential kimberlite targets.
• The bottom right figure shows two potential kimberlite targets marked with *.

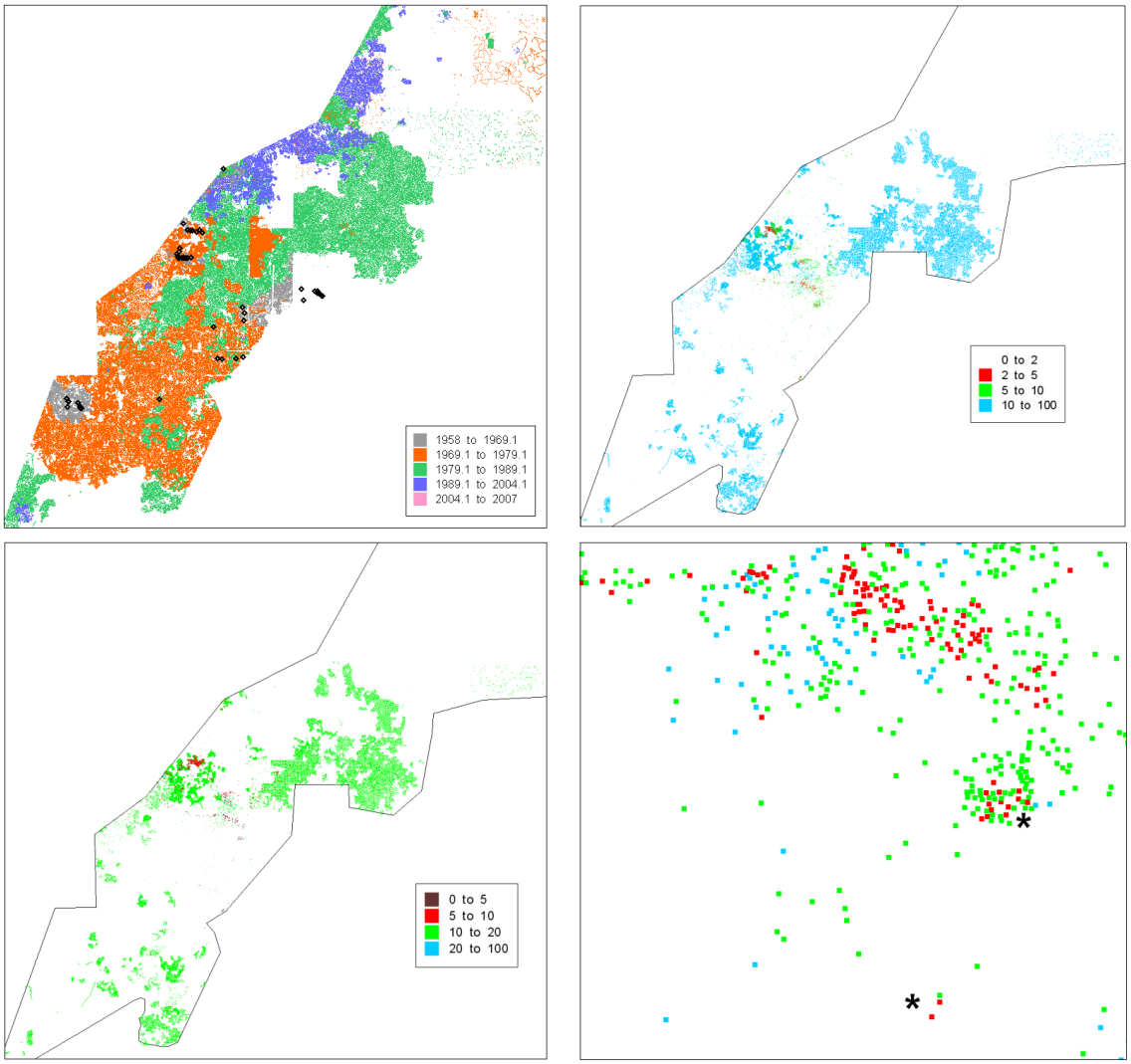


Figure 3.31: Application of a classification tree analysis on data from Northern Cape: South Africa

Where targets were identified, they will be checked against a database with other geophysical data and potential drill hole information to either prioritise, reject or target an anomaly for further investigation.

3.4.5 Final comment

Using classification trees is a sound statistical approach for interpreting indicator mineral data in the search for kimberlites.

However, a number of cautionary statements are worth noting:

• The model is dependent on the definition of the indicators. The indicator's distances are related to the level of erosion and the time/distance over which indicator minerals have been transported. Distance categories of 5, 10 & 20km will not necessarily apply equally well in all areas.

• The classification tree that is generated is dependent on the type of kimberlite used in the modelling. An ilmenite rich kimberlite will result in a different classification tree to an olivine rich kimberlite.

• This method **cannot** show the absence of kimberlites as it works on a training method that uses known kimberlites. Only kimberlites with a similar indicator mineral signature to those used for the building of the classification model will be identified.

• Each time a model is created, it should be placed in a library and run on all "unknown" areas in the database. Since most variables are collected in the same manner across the world, potential similarities could be found in the data from one project to the other.

• The application of these models must be done in conjunction with a geologist, as the results and input data are prone to errors in measurement, geological interpretation and judgement calls. A number of distance combinations and iterations of the classification tree building are required to reach a satisfactory solution.

This research has shown that classification trees can be used for the analysis of indicator mineral data and has demonstrated success in the case studies.

3.5 Expressing efficiency of geophysical flight line strategies

Once a kimberlite cluster has been identified and targeted for areal magnetic surveys an optimal flight path width and strategy can be followed to detect the actual kimberlite pipe(s). Different flight strategies can be tested on a known cluster of kimberlites where the size and number of kimberlites are known. This will enable an expression of probability to establish the most optimal flight path strategy.

On the Kaapvaal craton, the sizes of many of the kimberlite pipes are known and can be modelled at known location. An average size can be used for kimberlite pipes with unknown size. A method is required to express the probabilistic outcome for different flight strategies and a measure to express the most efficient strategy.

One approach to express the detection of kimberlite pipes is to overlay the area with flight paths represented as polygons (long thin rectangles) and circles representing the kimberlites and express through a polygon intersecting method, if any intersection occurred. Polygon intersection methods are readily available [69].

Figure 3.32 below shows a flight line passing through a kimberlite cluster. Depending on the requirements and sensitivity of detection, the outcome of the polygon intersection can express either:

1) The success in intersecting the kimberlite clusters, or

2) The probability of intersecting one or more of the kimberlites within a cluster.

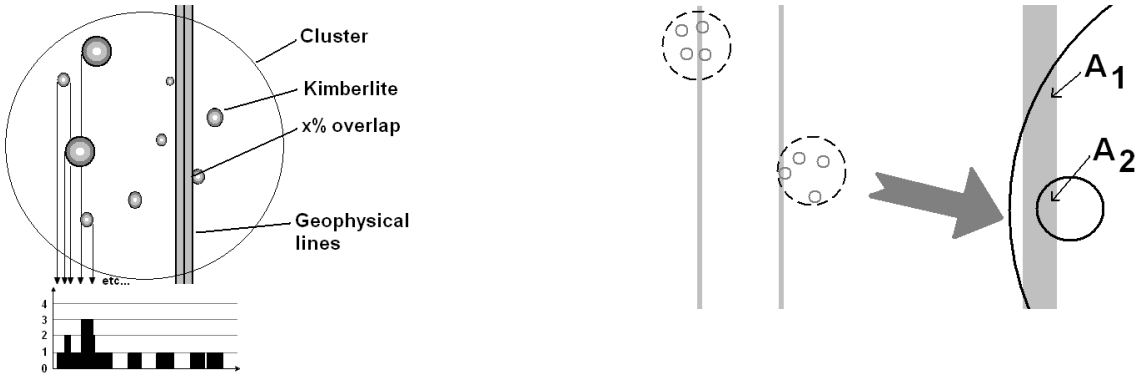


Figure 3.32: Simulating intersections of flight lines (A_1) with kimberlite pipes (A_2) and clusters

Multiple simulated flight paths can express a probability of intersection as:

$$\text{P(detection)} = \frac{\text{Number of times a kimberlite is detected}}{\text{Total number of attempts or total flight line length}}$$

Provided that the survey lines are vertical the likelihood of detection can be simplified and is proportional to the width of the kimberlites projected along the horizontal axis as demonstrated at the bottom of Figure 3.32(left).

Different flight path strategies can be considered and the most optimal chosen.

3.6 Optimal sample line spacing for indicator detection

[4] In the 18'th century, Georges-Louis Leclerc, Comte de Buffon posed the question to become the Buffon's needle problem (Appendix B):

> *Suppose we have a floor made of parallel strips of wood, each the same width, and we drop a needle onto the floor. What is the probability that the needle will lie across a line between two strips?*

The solution to this problem can be applied to optimise the distance between lines of exploration samples to detect indicator mineral trains.

With elongated lenses representing the indicator mineral trains, with random directions caused by local erosion processes, similarity can be drawn to the needle representing the indicator mineral train and the strips of the floor representing the distance between two exploration lines.

Although the deposition of minerals may not always be in a thin lens shape, an approximation can nevertheless help guide exploration on the spacing required between exploration lines.

The application of this method was used successfully in the exploration of marine gold in South America by Kleingeld & Prins [40] where the size of the lenses in which the gold was deposited

[4] Source: Wikipedia : http://en.wikipedia.org/wiki/Buffon's_needle: Accessed June 2009

was approximately known. In the search for kimberlites, the size of the indicator trains (based on the predictor variables identified from the classification tree building) must be determined.

Using the indicator minerals identified in the model building of the classification trees (J1, ilm-03, ilm-05) and interpreting them through a kriging process as a visual picture, gives some idea of the shape and size of the indicator mineral train on which a Buffon needle problem can be applied. Finding the optimal distance between exploration lines will ensure that in early stage exploration, a sample campaign has a line spacing of appropriate distance to express the probability of intersecting an indicator mineral train. Once an anomaly has been detected, infill sampling can be done.

Two examples of indicator mineral trains are shown:

Example 1: Images of an indicator mineral train in Canada showing the individual samples and interpreted data.

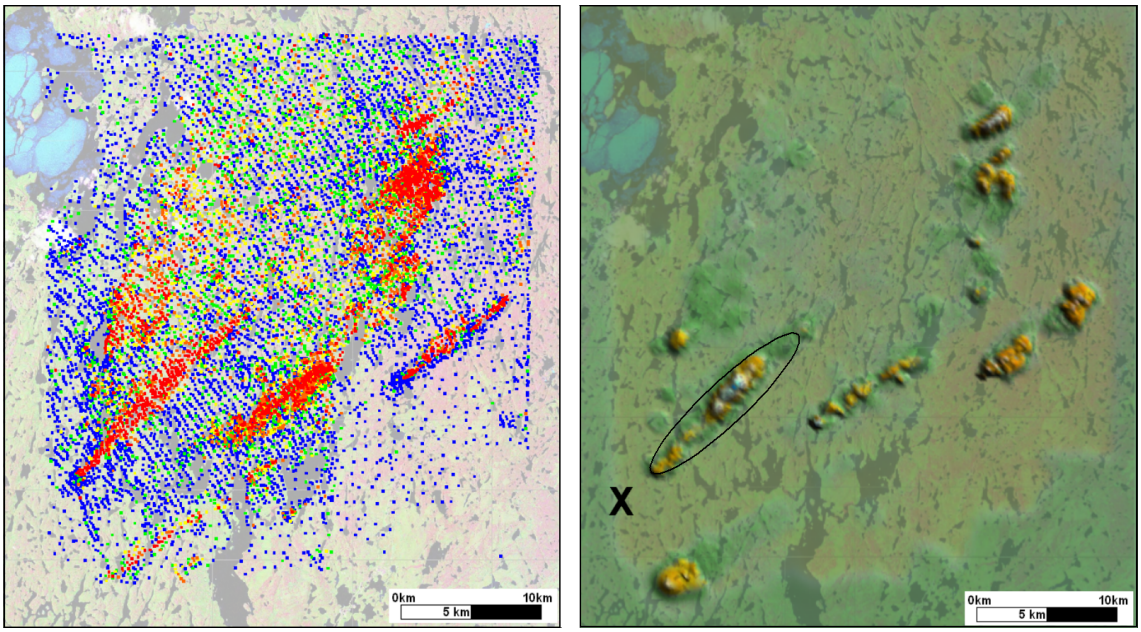


Figure 3.33: An example of an indicator mineral footprint

The warmer (red) colours show the higher counts of indicator minerals per sample. Interpolated with kriging and plotted as a shaded relief map (right), the higher relief (mountains) indicate the higher occurrence of indicator minerals. These are the indicator mineral trains searched for during the exploration process. Comparing against the scale on the map, the length of the indicator train can be determined for use in a Buffon's needle problem.

Example 2: Figure 3.34 shows the size and shape of an indicator mineral train based on the 0.3mm size fraction of Ilmenite data.

Interpolating the data with kriging, the results are plotted as a shaded relief map.

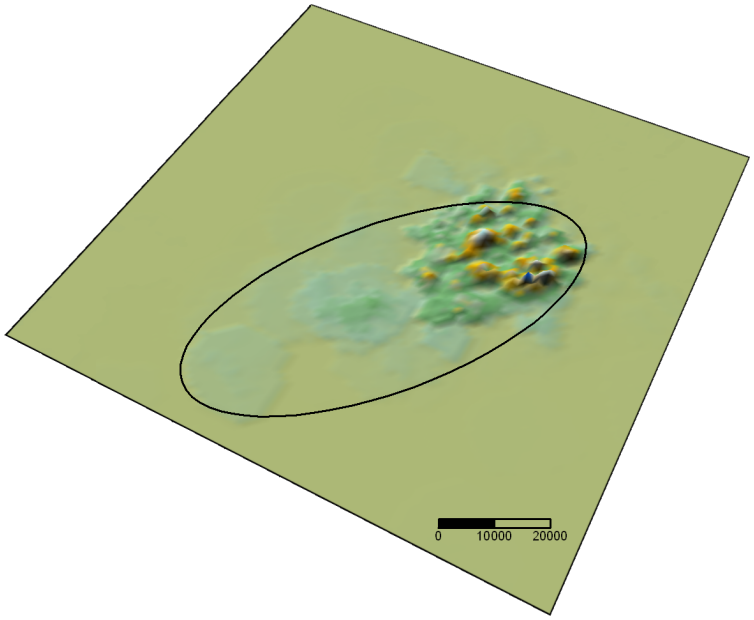


Figure 3.34: An example of an ilm-03 footprint

The higher relief indicates areas where ilmenite occur in abundance in the 0.3mm size fraction. Using the ellipse as a guide, the size of the indicator mineral train can be determined.

Applying Buffon problem to determine optimal distance between exploration lines

Appendix B describes Buffons problem in more detail and shows the solution to the needle problem as:

$$P(\text{needle touches a line}) = \frac{2d}{\pi a}$$

Interpreting the length of the indicator train as $2d$ by measuring the long axis of the ellipsoids from figures 3.33 and 3.34, with 95% probability of detection, the distance between two exploration lines ($2a$) can be expressed and determined as:

$$\text{Distance between two exploration lines} = \frac{4d}{0.95(\pi)}$$

In addition to this solution, research is ongoing to find the optimal sample shape, size and orientation to enable maximimum detection. A publication by Kleingeld, Lantuéjoul & Prins[41] researching this topic for the marine environment is detailed in appendix C, with relevance to the exploration of shapes of indicator mineral trains acting as entrapment features.

3.7 Concluding remarks

Exploration is very much a "hit and miss" process, but where some data is available, parameterisation of the exploration process will be beneficial.

The research detailed in this chapter demonstrated methods that can be applied to optimise sampling effort to be more efficient, as demonstrated with application in the case studies.

Chapter 4

Sample optimisation for grade when doing exploration in complex mineralisation environments

4.1 Introduction

When exploration projects are considered and sample optimisation is required for the first grade samples, the situation could be that geological information is available, with no or very little measurements for grade. Speculative mineralisation models and different sample scenarios are considered in an attempt to find the most optimal way to sample and collect the first grade data.

It is often difficult to determine whether one sampling campaign would be better than another and by how much, and it is desirable to be able to assess the relative performance of one campaign against the other.

Emphasis is placed on the fact that a sample optimisation study at this stage of a project is based on the potential for mineralisation and is done without any grade measurements, in other words, purely speculative. Only once the first sample campaign has been completed can definitive statements be made about the characteristics and measures of grade.

A method was researched by which non-grade information ("soft information") can be captured to express the potential of mineralisation for use in sample planning and optimisation studies of this speculative kind.

4.2 What is the question?

When simulation is used as a method to optimise sample campaigns, some grade information is usually already available that enables the building of models. With grade information not readily available at an early stage of exploration, an alternative approach is required.

If a methodology could be formulated that can give guidance (even relative between different strategies) to optimise sampling, benefit will be derived before embarking on expensive sampling campaigns.

The objective is that when no or very little grade information is available, insufficient to establish a spatial model and given that the only information available is an expectation of the mineralisation, a method is required to help rank the efficiency of different sample campaigns.

Once sufficient grade information has been accumulated to build a geological and spatial model and representing the underlying statistical models correctly, more robust and statistically sound stochastic simulation models must be applied for subsequent sample optimisation studies.

4.3 What is "soft" information?

"Soft" information for the purposes of this research is considered all the information that a geologist will consider in building a conceptual mineralisation model, including grade information when available. The combination of all of this "soft" information is assimilated into building the conceptual mineralisation model the geologist has.

The formation of more than one model is possible, as all interpretations can be analysed and used as different representations of reality, to provide a range of solutions.

Keeping in mind that the objective is to optimise the first sample campaign for grade determination and thus at this stage, no guarantee exists that the optimisation will be optimal, neither a guarantee whether mineralisation is actually present or not.

4.3.1 Examples of soft information

Different examples of "soft" information are shown below in figures 4.1 to 4.6. The examples do not relate to specific arguments around grade optimisation studies, but are shown as examples of geological interpretations that can be considered "soft" information.

Figure 4.1: In fluvial environments on the coast of Namibia, gullies act as entrapment features into which coarser gravel and diamonds accumulate.

If knowing that gullies are present, the size, depth and regularity thereof will provide an indication of potential mineralisation features (but no guarantee about the grade content).

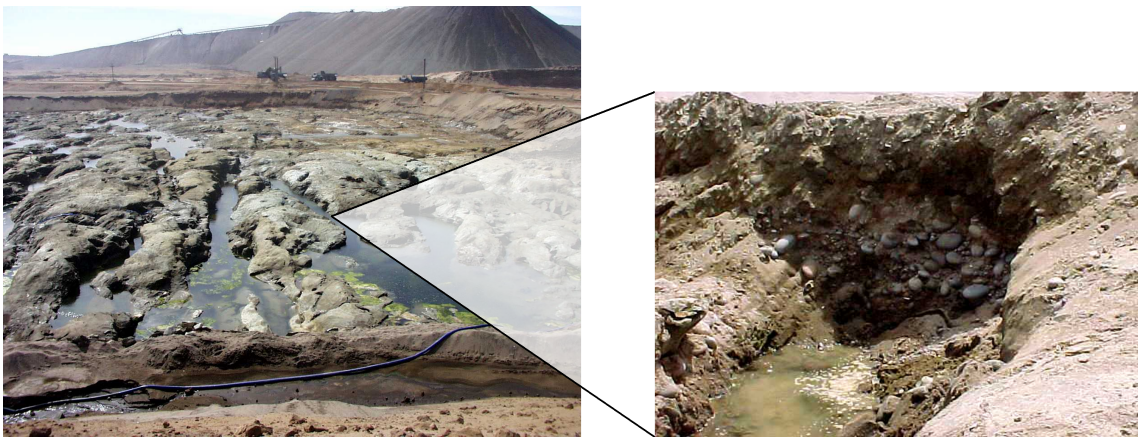


Figure 4.1: Gullies containing gravel and the potential for higher diamond content

Figure 4.2: Where entrapment features are fairly regular, similarity in mineralisation potential exists from one gully to the next. For example:

Figure 4.2: Regular gully pattern

Figure 4.3: Deep gullies, appropriately aligned relative to the direction of wave energy, will act as good trapsites. An abundance of gravel and higher grades are expected. An example of deep gullies is shown below.

Figure 4.3: Deep gullies with high mineralisation potential

Figure 4.4: Sub-marine environments are more challenging to interpret as no direct human interaction is possible with the in-situ deposit. Excluding the material recovered onboard the sample vessels, most observations and measurements are done with remote sensing techniques

that enables broad scale interpretations. Examples of interpreted "soft" information are shown.

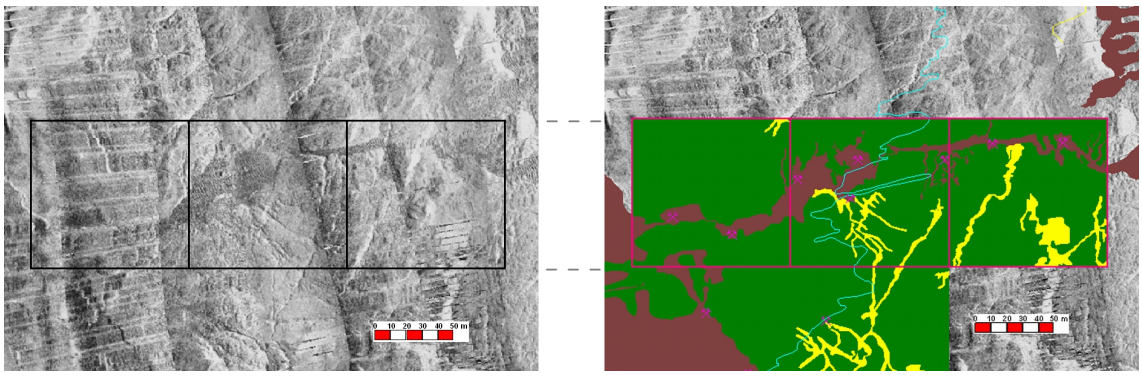


Figure 4.4: Side scan sonar of irregular gullied environment relative to 100x100m blocks(left), with geologist's interpretation in yellow (right) of potential mineralisation.

Figure 4.5: At some locations in the sub-marine environment, samples have indicated an increase in grade within gravel waves and although this argument cannot be generalised, it can provide "soft" information to assist in the model building process.

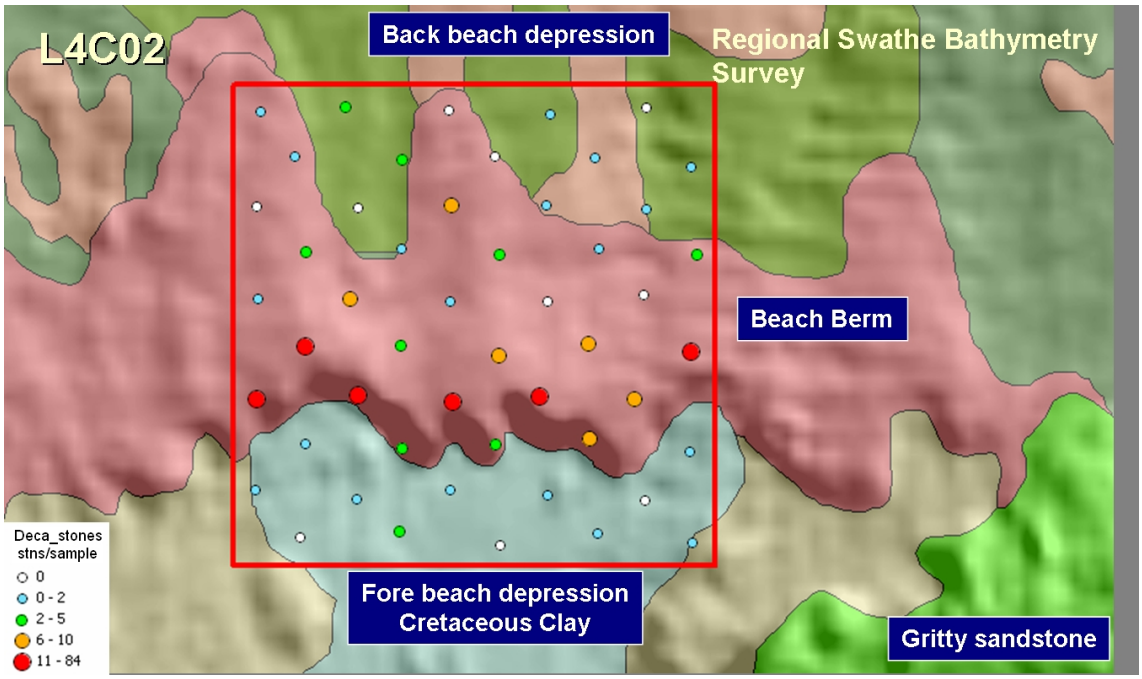


Figure 4.5: A beach berm showing an increase in stones per m^2 as the berm thickens

Figure 4.6: In some cases, visual observations, similar to sub-marine environments, can indicate where potential mineralisation could occur.

Figure 4.6: Interpretation of visual information indicative of mineralisation. A beach berm (left) and accumulated gravel (right)

The individual concepts and observations in figures 4.2 to 4.6 are not necessarily significant by themselves, but the collective interpretation is considered "soft" information that can be used and combined to build a conceptual mineralisation model. Experience in creating mineralisation models is gained with the sampling and mining of beaches and sub-marine deposits and the studying of different geological environments, through geological interpretation and parallels drawn with other similar exploration targets.

4.4 Capturing a conceptual mineralisation model

A method is required to capture and express the mineralisation model in such a way that it can be reviewed, criticised and improved upon, until an acceptable model is established.

A method by which this can be achieved is by creating a pencil sketch of the expected mineralisation, shading higher grade potential darker and the lower potential areas lighter.

Through the sketch, the scale of geological features is expressed, their relative spatial location to one another and through the grey scale intensity, an expectation of grade variation.

The sketch captures the sum total of all of knowledge, "soft" information and interpretation related to the exploration target and can be classified as an expert system.

According to Wikipedia [1]:

An expert system is software that attempts to provide an answer to a problem, or clarify uncertainties where normally one or more human experts would need to be consulted. Expert systems are most common in a specific problem domain, and is a traditional application and/or subfield of artificial intelligence. A wide variety of methods can be used to simulate the performance of the expert however common to most or all are:
1) the creation of a so-called "knowledge base" which uses some knowledge representation formalism to capture the Subject Matter Expert's (SME) knowledge and
2) a process of gathering that knowledge from the SME and codifying it according to the formalism, which is called knowledge engineering.
Expert systems may or may not have learning components but a third common element is

[1] Wikipedia website: http://en.wikipedia.org/wiki/Expert_systems, Accessed 7 October 2009.

that once the system is developed it is proven by being placed in the same real world problem solving situation as the human SME, typically as an aid to human workers or a supplement to some information system.

For example, consider the following gullied environment with entrapment features of various sizes, with a requirement to optimise a sample campaign. No grade data exists, but by using "soft" information an approximate mineralisation model can be created and sketched.

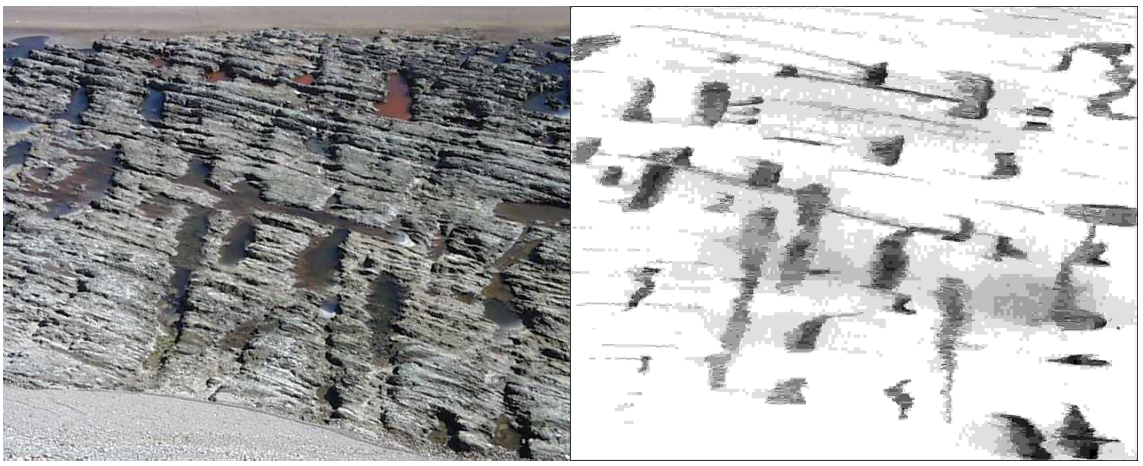


Figure 4.7: Creating a mineralisation model by sketching the diamond potential, taking cognisance of the need for geometric corrections

Once the mineralisation is captured through the sketching process, the image is scanned on a computer scanner and expressed as a grey scale image that can be manipulated statistically.

Typical questions that can be addressed using a simulation of this kind would be:

- If the area is covered by a sampling grid of 50x50 or 100x100m, how representative and variable would the estimates be?
- Would widely spaced trench sampling be better than regular grid samples?
- How different would the results be using a $2m^2$ sample versus a $7m^2$ sample?
- Small sample sizes may be required to penetrate the gullies, but then could be too small to measure diamond content accurately. If sample sizes are too small and many zero values are observed, how will this effect the accuracy of an estimate for the block?
- How different would a kriged panel result be from a weighted average?
- How could different grade distributions affect the decision making?

4.5 Manipulating the sketch

Once the sketch has been scanned, the image is exported as a text dataset containing pixels, each with an (ix, iy, grey_scale) [2] value.

[2]For bitmaps images, 0 = black (dark colour) and 255 = white (light colour). The sketch is created with darker colours representing higher grades. The grey scale is inverted during the export from bitmap image to (ix,iy,igrey_colour) data file to represent 0 = low grade and 255 = high grade.

The grey scale values range from 0 for low grade to 255 for high grade and forms a spatial dataset that can be statistically manipulated. This manipulation is on parameters related to the spatial structure (variogram) and grade distribution through the use of various transformations.

A series of manipulations are required to change the pixels data into a simulation and depending on the amount of grade information available, different steps will be followed.

4.6 Manipulating the sketch data into a simulation

Depending on the requirements and amount of grade data available, a list is given with a generalised flow of manipulation. The use of these steps will be demonstrated in the case study that follow.

Steps:

1. Create the hand sketch and export pixels into an (ix, iy, grey_scale) data file.
2. Regularise the sketch data to represent an equivalent sample support as would be observed and sampled in practice
   - Correlate the sketch with conditioning data (if available)
   - Extract data from a polygon covering the study area to create n values at sample support.
3. Normalise the sketch data by taking the sketch data with data limits $[0 \rightarrow \infty]$, sort the data in increasing order $z^{(1)}, z^{(2)}, z^{(3)}, \ldots, z^{(n)}$ and transform each of the n values as the k'th quantile of the standard normal cumulative distribution value $y^{(k)} = G^{-1}\left(\frac{k}{n+1}\right)$ to result in a Gaussian distribution dataset with limits $(-\infty \rightarrow \infty)$.
   - Calculate the variogram
   - Model the Nugget Effect
4. Take the conditioning data (if present) with data limits $[0 \rightarrow \infty]$ and as for the sketch data, transform the data to a Gaussian distribution $(-\infty \rightarrow \infty)$.
   - Calculate the variogram
   - Model the Nugget Effect
5. Due to the texture of the paper used for the sketch and the effect of regularising the sketch data, there will be no similarity between variogram parameters. The sketch data must be manipulated to reflect the same nugget effect as the observed data. If a nugget effect of $\triangle NE$ needs to be added:
   $NE_{required} = \sqrt{\triangle NE} \times g(0,1) + \sqrt{1 - \triangle NE} \times$ sketch data
   where $g(0,1)$ is a standard Gaussian random variable.
6. Back-transform the Gaussian sketch data using the conditioning data's normal score transformation table.
7. Seed the stones for use in the simulation exercise.

4.7 Case Study

A pocket beach on the west coast of Namibia was sampled with 39 samples and an indication in the confidence of the average grade is required.

Although the number of samples is limited, a spatial simulation of the deposit is proposed that could be sampled repeatedly, by locating samples randomly in proximity of the observed sample locations, to establish an indication of the variation in the estimate of the mean. Confidence limits must be determined on the repeated outcomes of the experiment and related to the observed mean. A conditional simulation is proposed for this study.

Target area and sample location plot

The target area sampled with 39 samples is shown in figure 4.8, with the map (top) showing the area with basic geological interpretation, outline of the target area and sample locations and (bottom) the spatial sample results, with sample locations showing the number of stones per $10m^2$ sample.

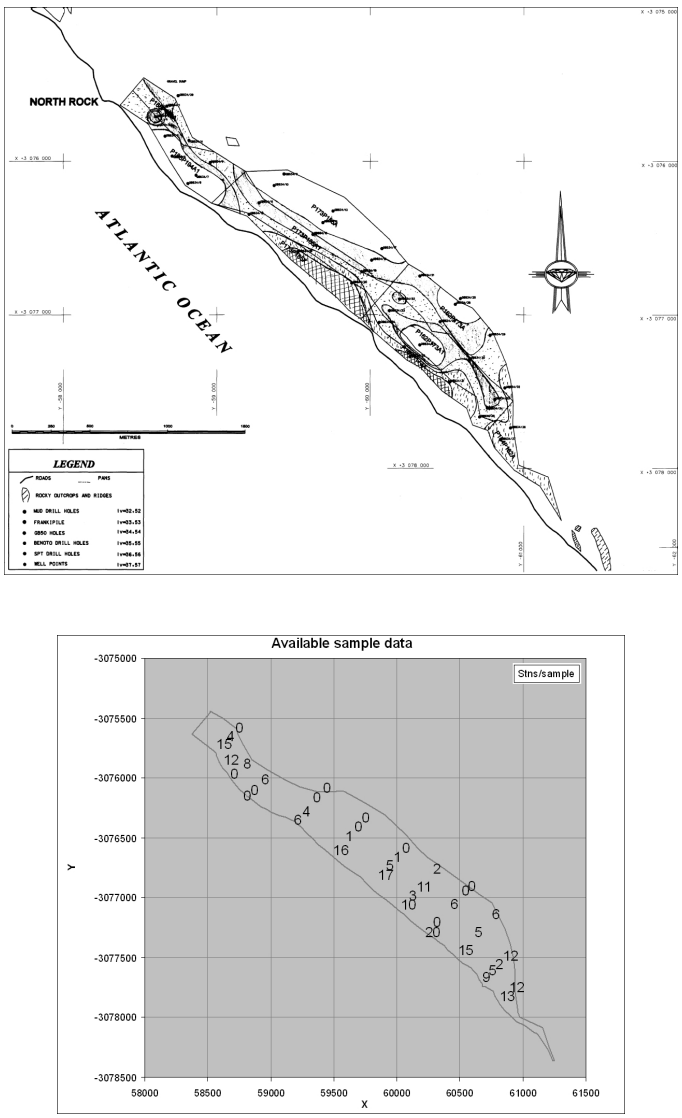


Figure 4.8: Target area sampled with thirty nine $10m^2$ samples

Sample data

Sample data had been collected along and across a pocket beach and the sample results, listed as stones per 10m^2 sample, are listed in the table below.

Table 4.1: Sample results

Stones per sample results for 39 samples (sorted by row in increasing order)												
0	0	0	0	0	0	0	0	0	0	0	0	1
1	2	2	3	4	4	5	5	5	6	6	6	6
8	9	10	11	12	12	12	13	15	15	16	17	20

with univariate statistics:

```
+------------------- U N I V A R I A T E    R E S U L T S ----------------------+
|                                                                                |
| Data File   : pocket beach.geo                                                 |
| Variable    : Stns                                                             |
| Minimum     :         0.0000                                                   |
| Maximum     :        20.0000                                                   |
+--------------------------------------------------------------------------------+
| # Observations  :        39                                                    |
| # Values = 0    :        12                                                    |
| Mean            :         5.7949      Minimum Value     :         0.0000     |
| Variance        :        34.7989      25th Percentile   :         0.0000     |
| Std. Deviation  :         5.8991      Median            :         5.0000     |
| %Coef. Variation:       101.7980      75th Percentile   :        11.0000     |
|                                       Maximum Value     :        20.0000     |
| Skewness        :         0.6927                                               |
| Kurtosis        :         2.2252                                               |
|                                                                                |
| 4 Largest values:       15.0000;       16.0000;       17.0000;       20.0000 |
| 4Smallest values:        0.0000;        0.0000;        0.0000;        0.0000 |
+--------------------------------------------------------------------------------+
```

The samples have been spaced closer to one another across the beach (across geology) than along the beach (along the expected geological continuity) and visual geological observations classifying the potential of mineralisation is available, based on the type and coarseness of material making up the pocket beach deposit. These visual observations, based on the characteristics of the gravels, acts as a proxy to indicate the continuity and degree of mineralisation.

The following points are relevant to the study:

a) A study is required to determine results for a specific question, related to variation around known sample locations and specific characteristics of the deposit
b) A generalised answer based on a large simulation domain would not suffice (the simulation requirement will not be adhering to ergodicity requirements)
c) A conditional simulation is the most appropriate approach to this study
d) Challenges regarding the limited number of available samples, size of simulation domain, accurate determination of the variogram and other statistics are recognised
e) The outcome of the conditional simulation and resultant study can be challenged due to the lack of sufficient numbers of samples.

Determination of variogram structure

For the conditional Cox simulation (Appendix F), the Gaussian anamorphosis and variogram parameters are based on only 39 samples and some parameters were determined with uncertainty due to the lack of sufficient sample data.

Two variograms are shown in figure 4.9:

(left) The variogram is of the untransformed 39 data points. The two main directions of anisotropy are shown along the beach (wider spaced samples: modelled with long range) and across the beach (closer spaced samples: modelled with short range). With the limited data available, some uncertainty is observed in the variogram structure determination. The closer spaced samples indicate a nugget effect of 53% of the total sill, whereas the structure along the beach indicate a nugget effect ratio of 18%. With more couples available at the shorter distances across the beach, the nugget effect of 53% is considered.

(right) The Gaussian variogram with anisotropy modelled according to parameters from the raw data.

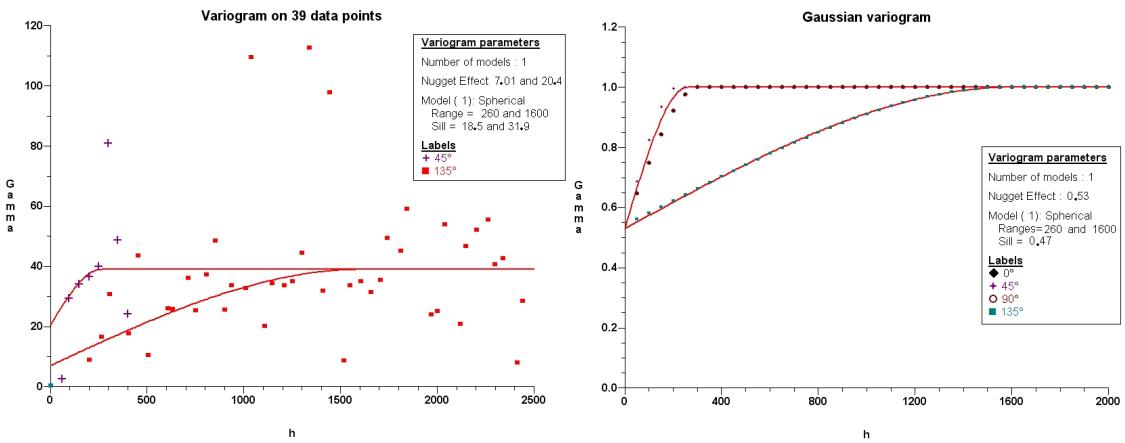


Figure 4.9: Variograms for observed data and Gaussian variogram used for the Cox simulation

A Cox simulation was created based on the parameters determined from a very limited dataset.

Result of the Cox simulation

Figure 4.10 shows the outcome of the conditional Cox process, highlighting the effect of the anisotropy modelled for the two principle directions.

The figure offset on the right in figure 4.10 shows the simulation after it was seeded with individual stones, with each black dot representing a stone on the beach.

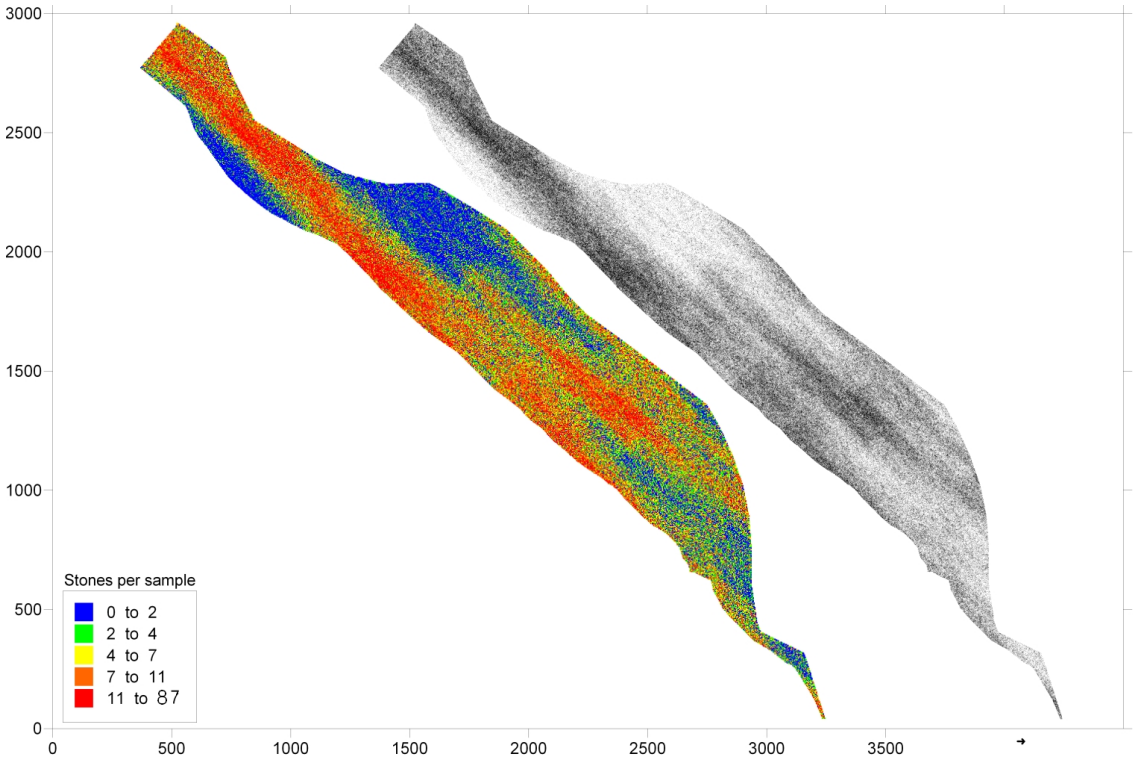


Figure 4.10: Outcome of the Cox simulation

Interpreting the simulation result visually, indicates too much continuity existing along the main direction of the beach (long axis of anisotropy) and generally it was felt that both the low and high grade areas of the deposit were extrapolated too far across the extent of the deposit (an effect of stationarity in the simulation technique's assumptions), not representing the expectation of the geologist.

A method by which the expectation of the geologist is captured through a pencil sketch is proposed.

Capturing "soft" information into a conceptual model and creating a sketch

Since the geologist had a good conceptual idea of the expected mineralisation, the geologist was asked to sketch his expectation of the mineralisation based on the sample data, field observations and general interpretation based on experience related to this and other similar beach deposits.

The result of the sketch was as follows:

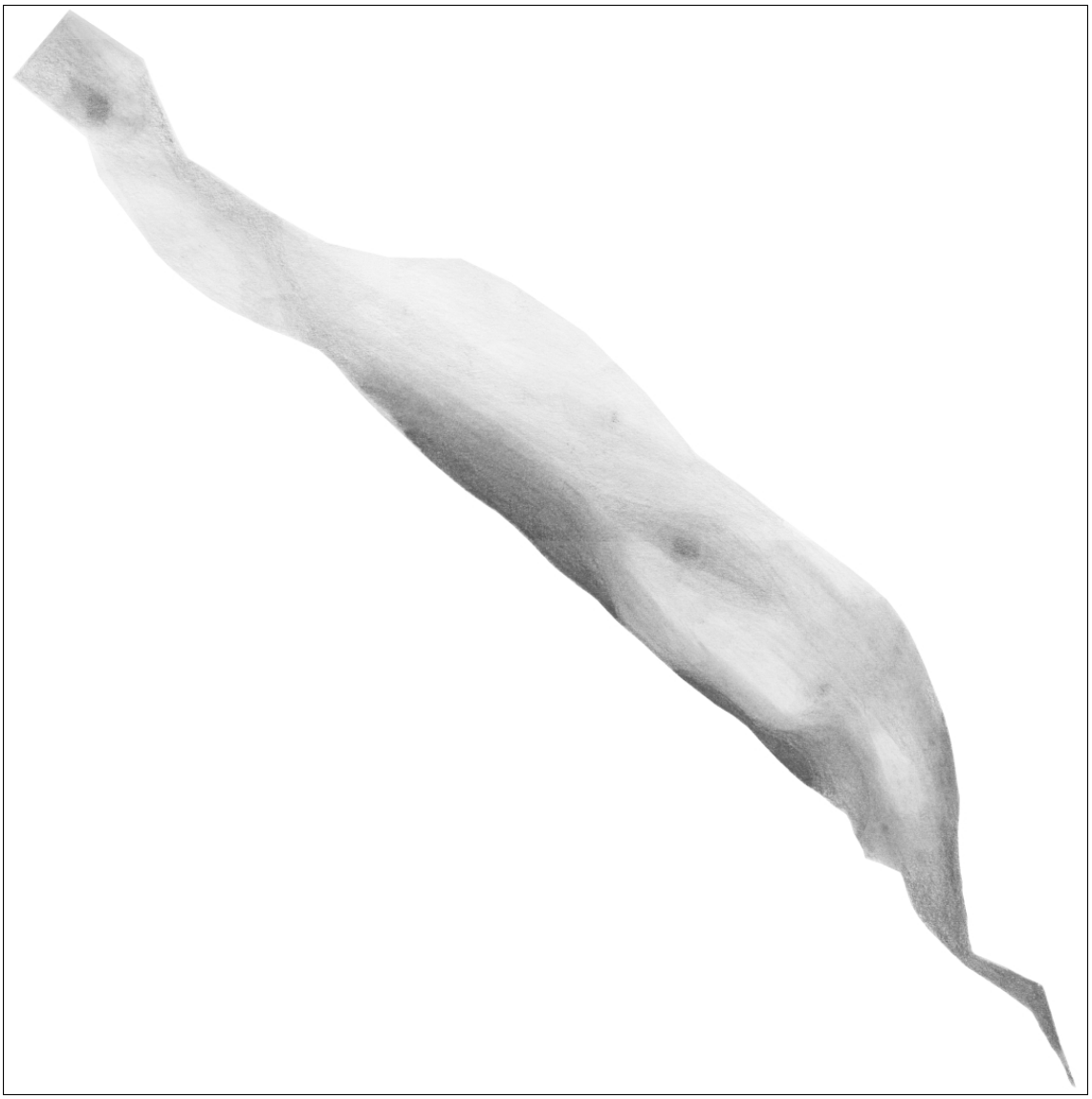


Figure 4.11: Geologist's interpretation of mineralisation potential

The sketch was scanned into a computer as a bitmap image (.BMP file). This file is a raster file with each pixel of the sketch represented by a grey scale intensity between 0 and 255, with zero representing a low mineralisation potential and 255 indicating a high mineralisation potential.

Correlating the sketch against observed data

This raster data from the sketch can be exported and manipulated as spatial information[3]. The software exporting the bitmap file relates each pixel from the sketch to a (X, Y) location relative to coordinate limits of the beach. Figure 4.12 shows the input screen exporting the

[3]Software was written by C. F. Prins to export raster bitmap files into a (X, Y, Grey scale intensity) format. Other formats such as Potable Greymap (from Paintshop Pro ®) can also be used.

.BMP file into a text file as spatial data, relative to coordinates with origin:
(X, Y) = (58 000, -3 078 400).

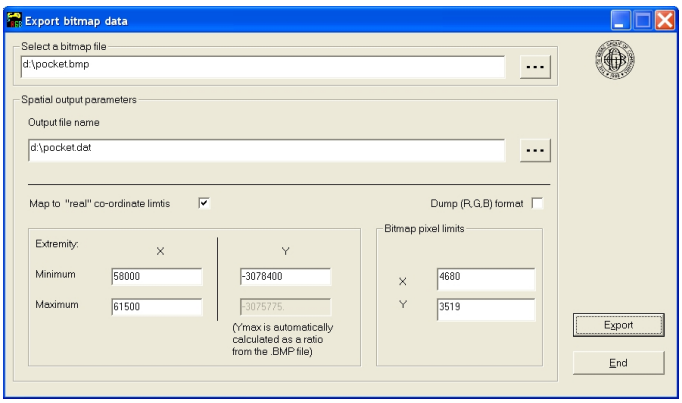


Figure 4.12: Software to export a .BMP file into (X,Y) coordinates

A test was made on the raster data to see if correlation exists between the grey scale intensity of the sketch and the 39 available data points. The correlation is plotted in figure 4.13, showing the relationship between the observed sample grades and the sketched grey scale pixels (correlating data against pixels up to a circular distance of 20m away). The result shows similarity and was accepted for further use.

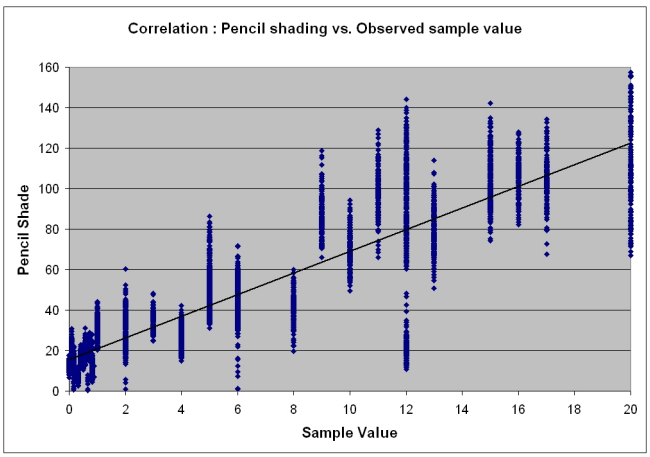


Figure 4.13: Correlation between sketched mineralisation (grey scale) and available grade information

Regularising pixels to represent the equivalence of the observed sample size

The sketched domain contains 4680x3519 pixels which, when related to the equivalent extent of the target area, is equivalent to samples of 0.73m^2 in size. The 39 available data points are 10m^2 in size, which requires the bitmap to be regularised into equivalent sample sizes of 3.16mx3.16m (area=10m^2). Regularising the sketch data resulted in an output file containing 1 190 700 pixels, each equivalent to a 10m^2 sample. This regularisation process aligns the

number of pixels in the bitmap figure, to the equivalent number of samples that would cover the full extent of the exploration target area (a union of samples forming the target area).

After regularising the grey scale data to form the pixels, each pixels represents the equivalent of a $10m^2$ sample.

This newly formed dataset is further manipulated to align spatial and grade characteristics with the observed data.

The sample plot of the regularised grey scale image (now on $10m^2$ sample support) is shown in Figure 4.14.

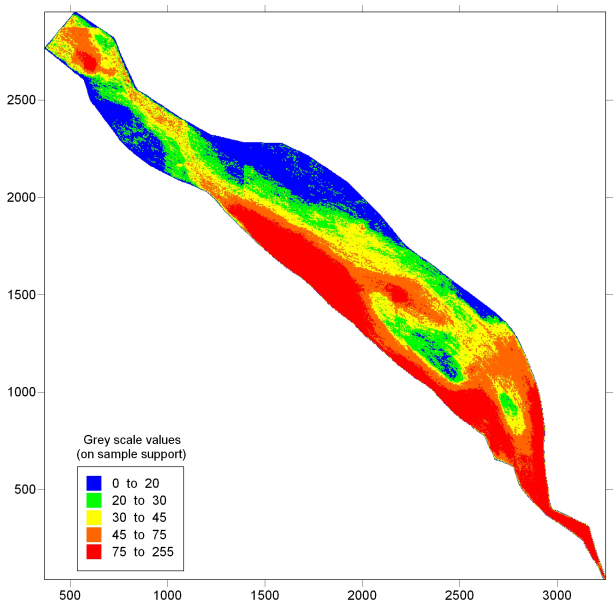


Figure 4.14: Regularised sketch, with each pixel equivalent to $10m^2$

Manipulating spatial characteristics of the sketch

With the grey scale data regularised the following can be noted:

a) There is no expectation that the variogram structures and ranges obtained from the sketch will relate to the structure(s) of the original sample data
b) The nugget effect to total sill ratio of the variogram of the observed data is expected to differ from the ratio observed from the regularised sketch. The texture of the paper the sketch was drawn on, as well as the effect of the smoothing caused by the regularisation process affects the spatial behaviour of the sketch.
c) The simulation study is required to investigate a sensitivity measure related to variation at specific sample locations. This variation is closely linked to the behaviour of the nugget effect and first few metres of the variogram of the data. If the 39 data points can guide the behaviour of the nugget effect and variogram at short distances of the sketch, this must be incorporated into the simulation outcome.

The nugget effect of the spatial structure can easily be changed if the data is transformed into Gaussian space.

The nugget effect of the variogram of the regularised sketch data (transformed to Gaussian space), is modelled as 4% of the total sill, whereas the nugget effect ratio of the observed data is 53%. The nugget effect is modified with the addition of a 0.49 factor to reflect the observed nugget effect by using the transformation:

$$NE^*_{\text{simulation}} = \sqrt{0.49} * g(0,1) + \sqrt{1-0.49} * \text{sketch data}_{gaussian}$$

where $g(0,1)$ is a standard Gaussian random number. The above results as: The variogram of the regularised sketch data is shown in Figure 4.15.

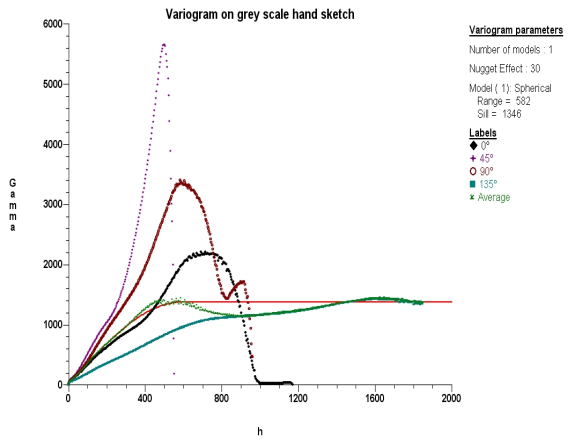


Figure 4.15: Variogram of regularised sketch data

The regularised sketch data (on 10m^2 support) is transformed into Gaussian space and the variogram calculated from this data is shown in Figure 4.16. The two main directions of continuity (along the beach and across the beach) are shown in Figure 4.16 (right), with the shorter range across the beach and the longer range along the beach.

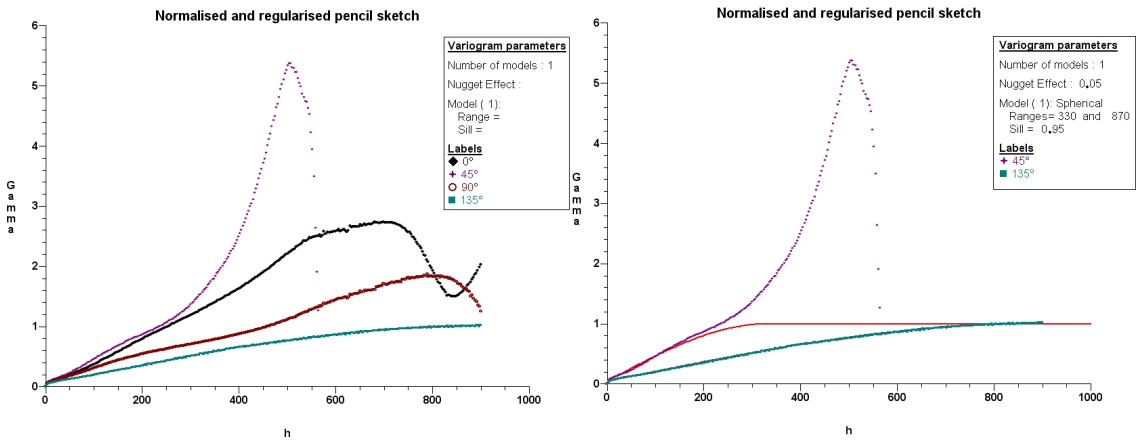


Figure 4.16: Variograms for regularised Gaussian sketch data

Once the Gaussian random component has been added to the nugget effect, the nugget effect to total sill ratio is increased to 53%, equivalent to the ratio of the 39 observed data points.

The variogram calculated on the modified sketch data is shown in Figure 4.17, which reflects the 53% nugget effect to total sill ratio.

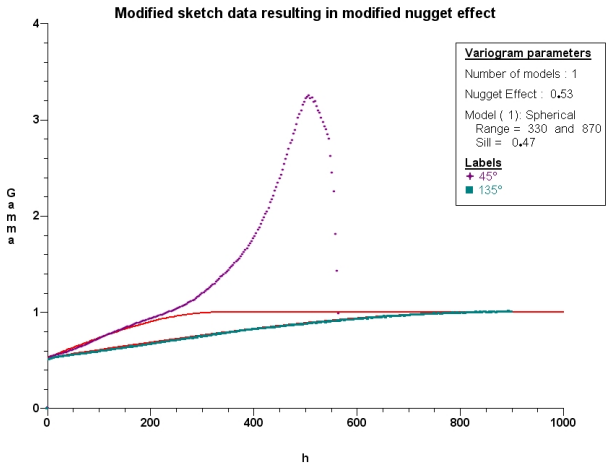


Figure 4.17: Experimental semi-variogram showing 53% nugget effect and models fitted to the two main anisotropy axes

The sketch data now represent:
a) The geologist's best interpretation of the spatiality of the mineralisation
b) Is on a 10m^2 sample support so that manipulation of the sketch data is on the same support as the observed data
c) As Gaussian data, the sketch data has the equivalent 53% nugget effect to total sill ratio as the 39 observed data points.

A last step in the process is for the Gaussian data to be back-transformed to reflect the histogram and statistical properties of the original untransformed 39 data points.

Transforming and back-transforming the data

The 39 observed data points are used to construct a normal score transformation table which is used to back-transform the sketch's Gaussian values to stones/10m^2 sample values. Due to the number of zero values observed in the data, a small random number was added to the zero values in order to smooth the origin of the graph and stabilise the transformation process.

Schematically:

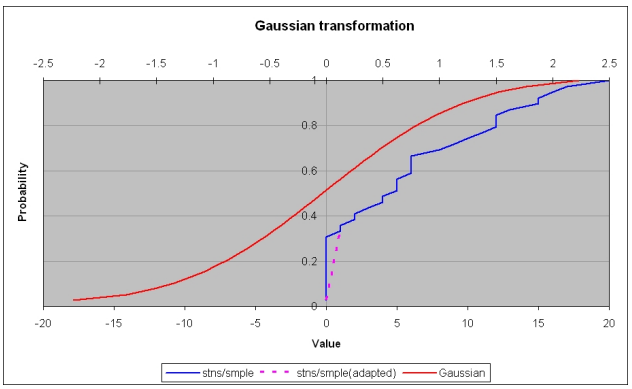


Figure 4.18: Normal score transformation of sample data

A back transformation is applied to the Gaussian values of the sketch data using the normal score table determined from the observed data. A check was done on the variography of the data after the transformation was done, to confirm that the 53% nugget effect to total sill ratio is maintained. The variogram is:

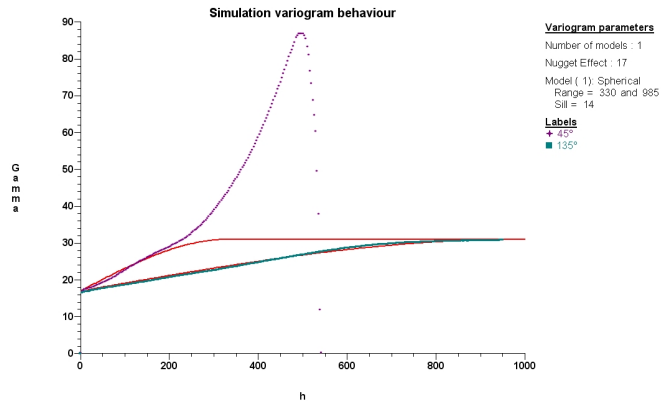


Figure 4.19: Variogram on back-transformed data showing 53% nugget effect

The results were as expected and shows a 53% nugget effect to total sill ratio.

Seeding and creating the simulation for analysis

Each data point of the sketch data, after all the manipulations were done, represents a sample of $10m^2$ with univariate statistics equal to the original 39 data points, as well as expressing the spatial behaviour close to zero distances (the nugget effect) similar in behaviour as that observed in the 39 original data points.

This is the end result of the manipulation and form the output result of the simulation process.

Referring back to Figure 4.9, two possibilities for the nugget effect ratio were observed, and the exercise above was repeated to create a second simulation with a nugget effect ratio of 18%.

Both simulations were seeded with stones and the results are shown in Figure 4.20, where each black dot represents a diamond. The two simulations are plotted side by side for illustrative purposes.

An analysis with repeated sampling was done whereby the sample locations are randomly located in close proximity to the locations of the 39 original samples and the results are analysed.

Analysing the two simulations

Each of the 39 sample locations from the observed dataset was used as a starting location and a random location was chosen in a surrounding ellipsoidal area, with axis lengths of 100m (along beach) and 50m (across beach). At each randomly chosen sample location, the stones

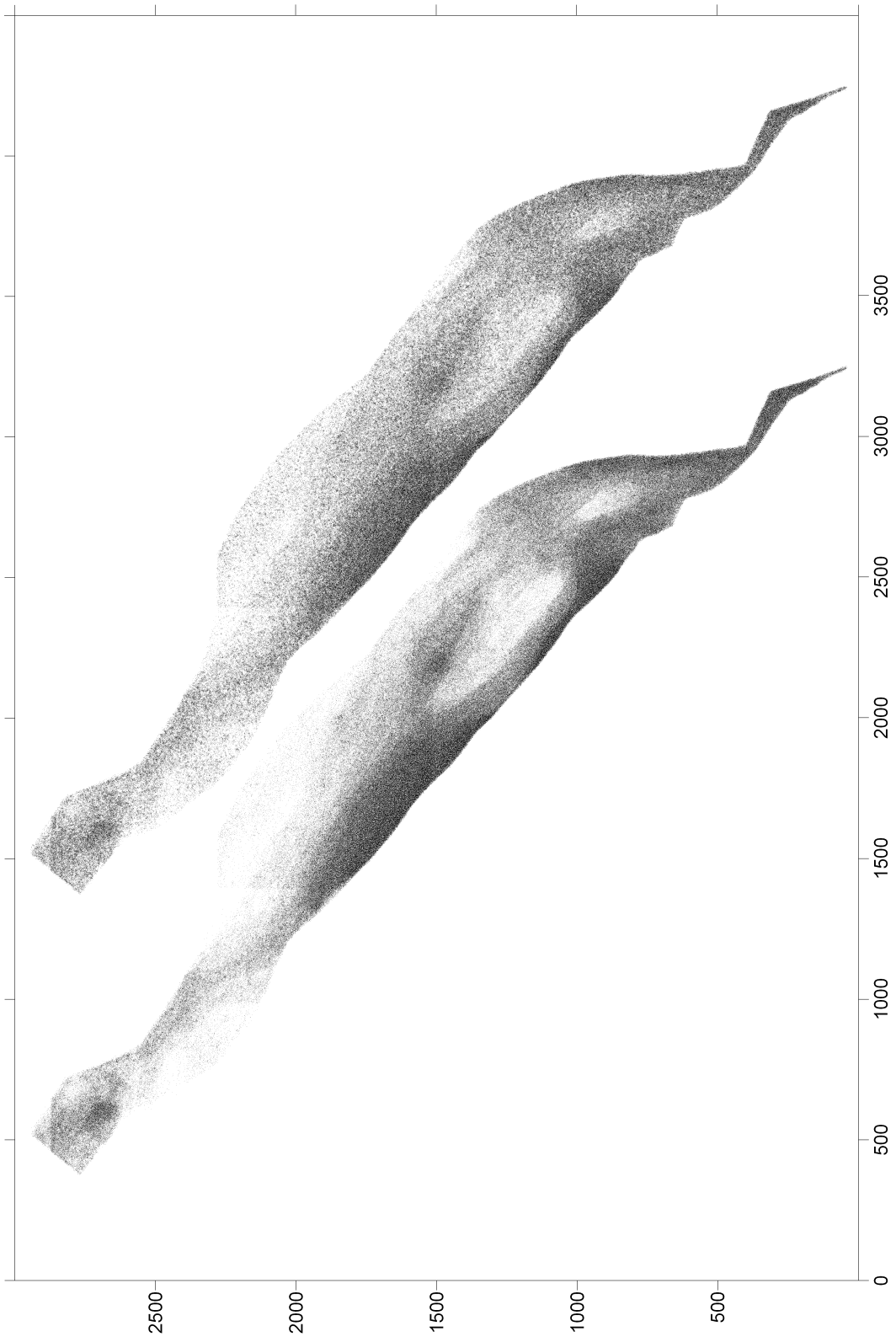


Figure 4.20: Simulations of stones for 53% and 18% nugget effect models

were recovered for a 10m^2 sample. One hundred repetitions of the sample campaign were done, the average calculated in each case and the results compared to the actual grade of the simulation. Results are shown in Figure 4.21.

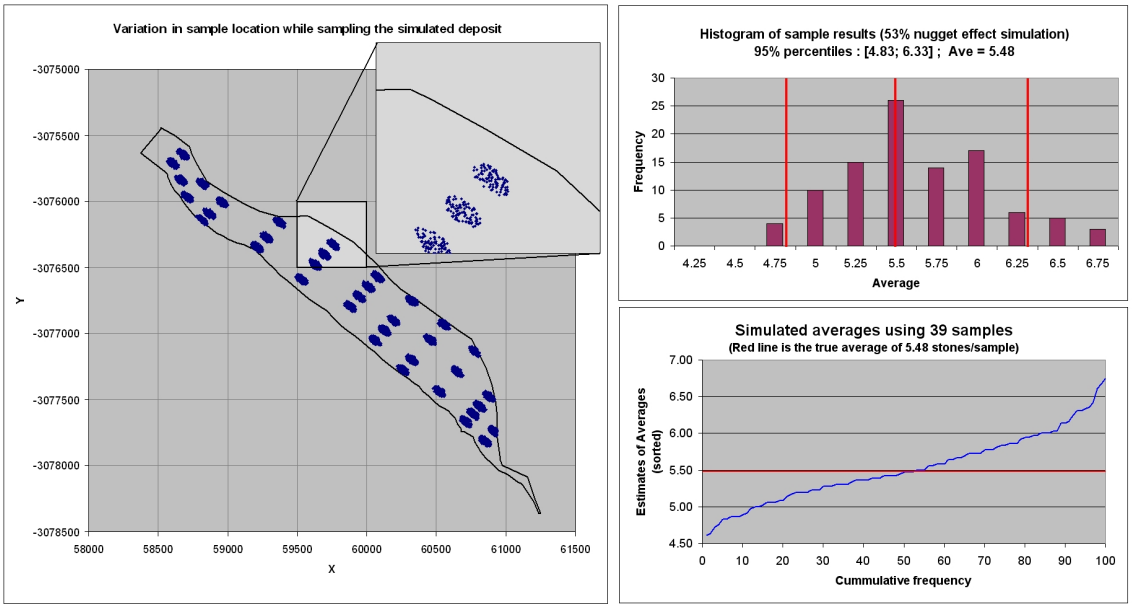


Figure 4.21: Analysis of sample variation on 53% nugget effect simulation

A second evaluation was done whereby increasing numbers of samples were randomly located inside the exploration target area and the mean estimated. This was repeated 100 times on both simulations (53% and 18% nugget effect simulations), estimates and percentiles of the mean recorded and plotted as trumpet curves relative to the simulation grade. Results are:

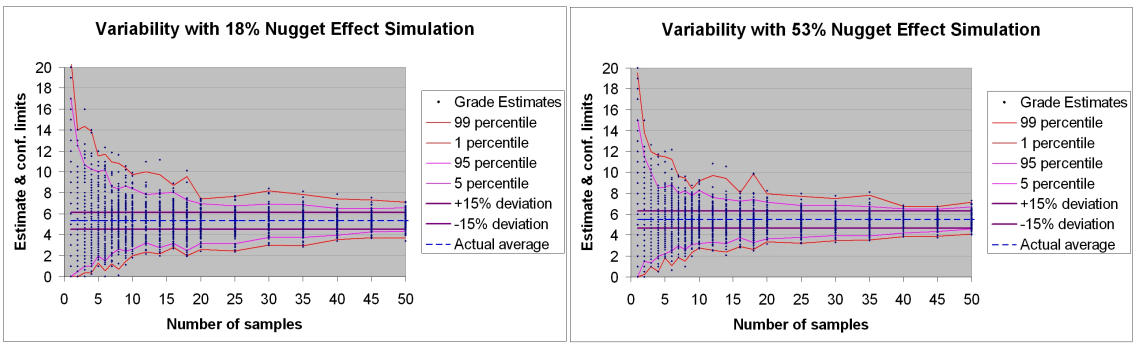


Figure 4.22: Analysis of number of samples relative to mean grade

Result

Interpretation of the results shows that the 5% and 95% percentiles are respectively 13% and 15% different from the true grade of the simulation and the trumpet curves have stabilised to reduce only marginally by the time 39 samples were taken.

As a concluding result, the 39 samples are considered representative to express the true grade of the exploration target within 15% accuracy.

4.8 Discussion

This approach to simulation can be criticised for a lack of having more descriptive statistical models and the process of creating the simulation is closer related to a series of data manipulations. This is fully acknowledged.

Examples of shortcomings are: This simulation method makes use of an expert system that is not repeatable and on different days the same geologist will shade the images differently; The skill and experience of the geologist plays a vital role as it is the interpretation put into the sketch that must adequately capture the mineralisation model; There is no control over the type of variogram model being sketched - for most of the test work done, the sketches showed a tendency to exhibit spherical variogram models characteristics, unlike data, that may have other model types; Multiple realisations of the same sketch is not possible.

Acknowledging the shortcomings of this approach, this method attempts to add specific value to the decision making to, without any data available and in geologically complex environments, attempt to provide a "more-correct-way-to-proceed" solution in providing guidance very early on in an exploration project when no grade data is available, to help determine the most appropriate sampling solution.

Once some data have been collected, the use of stochastic spatial simulation models must immediately be applied, as they are model driven and statistically robust. This is also the case when larger stationary deposits needs to be simulated, where non-conditional simulations would be the correct approach.

However, at the preliminary sampling stage (referring back to Figure 1 and the area of research for this thesis), no definitive statements about the expected variogram or statistics of the data, that is still to be collected, can be made.

For the purpose of optimising the first grade samples of a sample campaign, not having any grade information available and making use of speculative parameters in a complex geological environment, this sketching simulation can be used to compare one sample strategy against another.

Chapter 5

Sample optimisation in marine placer environments

5.1 Introduction

When De Beers Marine (Pty.) Ltd changed it's focus from an exploration to a mining company, a large portion of the mineral resource estimates had been estimated using results obtained from a sample tool measuring 0.72m^2 and a sample spacing of 100x33.3m. Mining operations commenced with test mining in four geologically different environments that included 25x25m blocks estimates and it soon became evident from recoveries that the small sample tool was somehow related to poor production recoveries. It was especially in the high grade areas where the most significant differences were recorded. There were several possibilities for the perceived poor recoveries, one of which was the impact of the representivity of sample results from the small sample tool.

Problems were identified with the mining process (in terms of efficiency) and improvements were made to the treatment and recovery systems. The investigation into recovery inefficiency ultimately focussed on the representivity of the 0.72m^2 (Mega-drill) sample tool.

The representivity of samples and the reduction of dispersion variance through the use of larger samples was discussed between W. Kleingeld & G. Matheron and as a result, simulation technology was developed by Kleingeld & Lantuéjoul [39] to simulate marine and beach diamond deposits. Creating suitable simulated deposits enabled different sample sizes and spacing to be tested with a view to improving estimation.

The 0.72m^2 was identified as a potential problem in terms of representivity and a project was initiated to investigate and establish the optimum sample size necessary for use in the marine environment. This tool needs to be used to sample in different geological terrains with varying grades and high integrity.

Since the introduction of a new sampling tool required significant capital to purchase a vessel and build a plant, rigorous assurance and scientific reasoning were required to establish an optimal size for the tool with a high level of certainty.

Geostatistical simulation making use of the Cox process by Kleingeld et al. [39, 42, 43] was applied and the 0.72m^2 Mega-drill data was incorporated.

5.2 Development of the Cox process

There are a large number of barren samples (i.e. no diamonds recovered) in the Mega-drill samples and an associated high nugget effect is apparent from the variogram. Conventional simulation models are not sufficiently suitable for use in sample optimisation studies of this nature.

Kleingeld and Lantuéjoul [39] developed the first Cox simulation methodology, which was later enhanced to allow the incorporation of conditional points as described in Kleingeld, Lantuéjoul, Thurston & Prins [43] in 1997. A description of the conditional Cox process is detailed in appendix F.

5.3 Case study

5.3.1 Background

The first application of the Cox process in a marine environment was aimed at establishing an optimal size for a sample tool for use in the Atlantic I mining area off the coast of Namibia by Duggan [20] and Prins [54, 53]. The early simulations were nonconditional and the information available for creating the simulation models was:

a) The 0.72m^2 samples and
b) Production data on a 50x50m block scale.

The accuracy of both datasets was problematic but it was assumed that the production results were more reliable. Given the observed differences between estimates and production results, it was assumed that the samples were nonrepresentative and the source of error.

To enable the creation of a simulation, the statistics on a sample scale were manipulated to enable meaningful comparisons to be made between the input parameters (based on nonrepresentative samples), the simulation outcome and the production data. The manipulation is graphically summarised in chapter 2 (Figure 2.8).

The stone distribution of the sample data was modelled with a flexible distribution using two parameters (ν and p) by a Negative Binomial Distribution (NBD).

The Negative Binomial Distribution, denoted as NBD (ν,p), as described by Evans [22] and Kleingeld [37] has probability mass function

$$P(X = x) = \frac{\Gamma(\nu + x)}{\Gamma(\nu)\ x!}\ p^x\ q^\nu \quad \text{with parameters } \nu > 0,\ 0 < p < 1,\ x = 0, 1, 2, 3, ...$$

with $q = 1 - p$ and with the first two moments:

$$mean = \frac{\nu p}{q}$$
$$variance = \frac{\nu p}{q^2}$$

Suppose that each stone is recovered with probability α and not being detected with probability $\beta = 1 - \alpha$. The total number of recovered stones follows a NBD $\left(\nu, \frac{\alpha p}{q + \alpha p}\right)$

After some manipulation, it can be shown[1] that $\alpha = \frac{\text{observed mean}}{\text{actual mean}}$.

After this scaling, the mean grade of the simulation compares favourably with the mean grade observed in the production data.

Analysis[2] based on the simulation shows that the optimal sample size for sampling the Atlantic I deposit is approximately 10m^2. After the 10m^2 sample tool had been in use for a few years, a need was identified to verify the results of the first study by also incorporating the 10m^2 sample results collected to date. The aim was to determine the sensitivity of using a tool smaller than 10m^2 , because the intention was to utilise a smaller vessel that could not deploy the larger tool.

A conditional Cox simulation of stones per sample was created for the follow-up study.

5.3.2 Problem definition

De Beers Marine was considering converting a smaller vessel (smaller than the current sampling vessel) to accommodate a tool larger than 0.72m^2 , but smaller than 10m^2 , to enable more cost efficient sampling.

Different configurations of sampling tools were considered, but the choice was constrained by engineering restrictions, mainly related to size and weight limitations of the vessel. Two of the preliminary sample tool configurations that were considered more suitable were:

a) A large diameter vibracore-type drill with approximate diameter of 2m (area = 3.14m^2)
b) A Wirth drill (which will require a drill tower) with diameter around 3m (area = 7m^2).

A study was undertaken to reassess the results from the previous study and to firm up on a minimum sample support size required for representative sampling, taking the 3.14m^2 and 7m^2 sample sizes into consideration for engineering onto the sample vessel.

5.3.3 Methodology and data selection

In some mining areas, a combination of both 10m^2 sample results and production data (50x50m blocks) are available. These areas occur in different geological environments and were identified as potential study areas.

There was sufficient numbers of samples for the accurate determination of the parameters required for the Cox process and a conditional Cox simulation was done. Comparisons were made between the conditional simulation and production data as part of verifying the simulation.

Sample results were sourced for five different geological environments and the most complex environment to simulate and analyse was selected as the study area based on the following criteria:

a) The highest nugget effect to sill ratio for the variogram
b) The shortest variogram range,
c) The lowest grade

[1] *Simulating marine deposits using the Cox process* by C F Prins, De Beers Internal Report, February 1995, Appendix A

[2] *Geostatistical Evaluation of a marine deposit by S Duggan, WITS M.Sc project report, 1995*

d) Potentially the most zero observations (highest proportion of zero values) in the sample results.

The most complex area was selected as results determined on such a study would also be applicable to the spatially less complex and higher grade areas.

One such an area was identified for the analysis and a study was done based on the dataset provided.

5.3.4 Variography

For a study of sample sizes smaller than the scale on which the conditioning data is available, the variogram must be deconvoluted to the smallest support and used with modified sample statistics to create the simulation on point support.

Two approaches to determine the appropriate parameters for the simulation were followed:

5.3.4.1 Deconvoluting the variogram using the NBD distribution

A method to scale the mean and variance on different sample support sizes, using the NBD distribution was considered.

By expressing the moments of the NBD as:

$$mean = \frac{\nu p}{q}$$
$$variance = \frac{\nu p}{q^2} = \frac{mean}{q}$$

the linear relationship between the mean and variance with a slope factor of $\frac{1}{q}$ is exploited. If the sample data is progressively regularised into blocks of 100x100m, 200x200m, 250x250m blocks and successive mean/variance ratios are calculated, it was intended to establish if an extrapolation could be done to establish the mean/variance relationship for samples less than 10m^2. This approach did not produce satisfactory results.

5.3.4.2 Deconvoluting the variogram by iterating variogram parameters

This approach was tested by applying successive iterations of parameters for the variogram for which the relationship $\gamma_v(h) = \gamma(h) - \gamma(v, v)$ holds.

Due to the long variogram range (relative to the size of the sample) and high nugget effect, no acceptable results could be obtained.

A method to successfully deconvolute the variogram was not found and a decision was made to create the simulation using the parameters derived from the 10m^2 samples. This creation of the simulation is not theoretically correct and rigorous validation on the outcome of the simulation was done to ensure that a study on samples smaller than 10m^2 would have some validity. An alternative to this course of action could not be found.

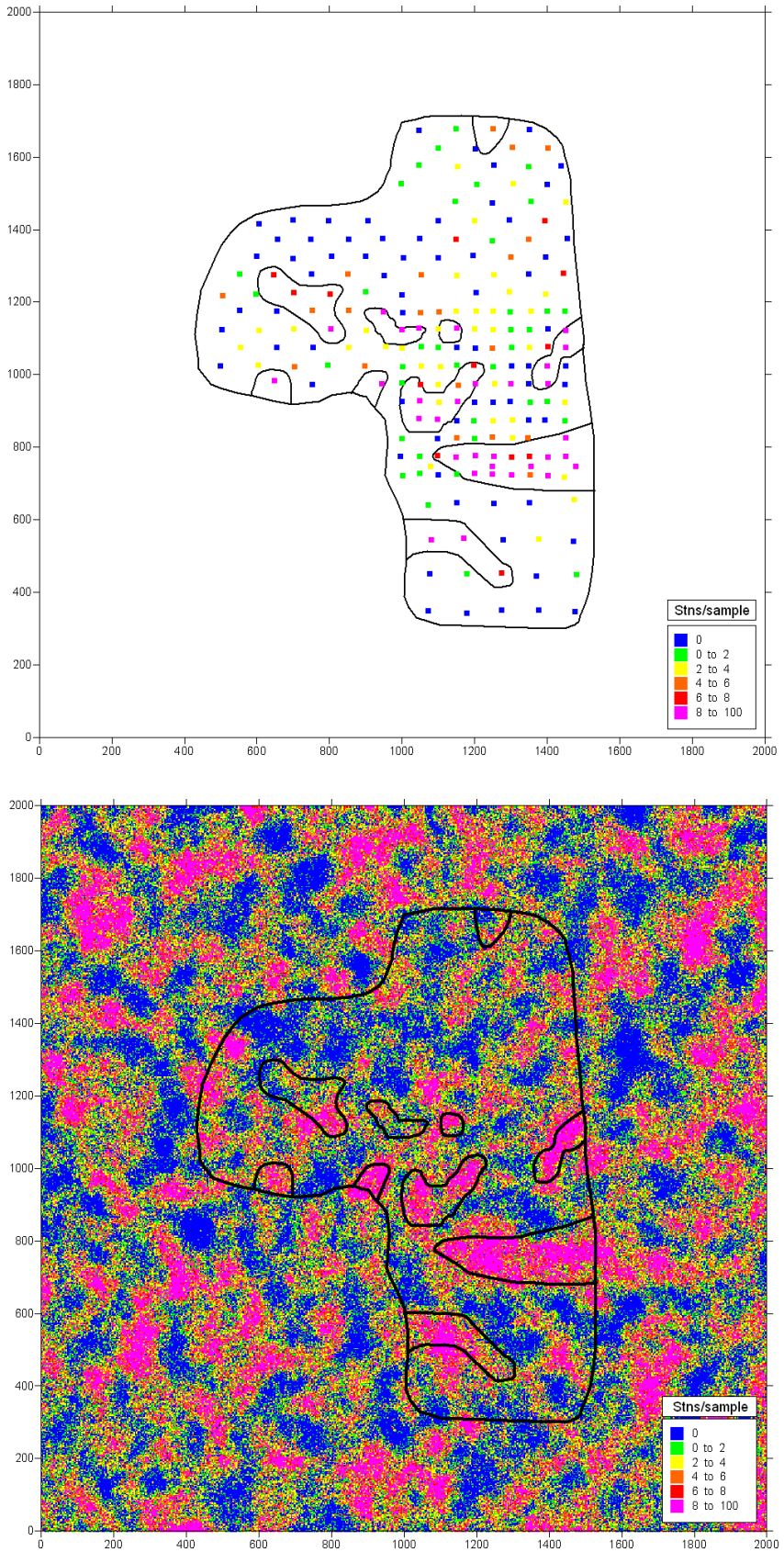


Figure 5.1: Conditioning data (top) and conditional Cox realisation (bottom)

5.3.5 Simulation and validation

A single realisation of a conditional Cox simulation was created as shown in Figure 5.1. The simulation domain encompasses an area of 2x2km containing conditioning data and the high grade areas are highlighted for reference. The simulation domain is roughly 20 times the variogram range, with warmer colours representing higher sample grade values (see Figure 5.1 bottom). A series of tests were done to validate the outcome of the simulation.

Validation 1:

A comparison between sample data on $10m^2$ sample support and the simulation results on the same support is given in the table below.

Table 1: Comparison of statistics between $10m^2$ data and simulation outcome

Statistic	Sample data	Simulation
Mean	3.49	3.79
Variance	22.44	21.85
Zero proportion	0.33	0.315
Variogram Range	90	97
Variogram Sill	11.8	11.2
Variogram Nugget Effect	10.6	10.4

The similarity between the variography parameters and univariate statistics indicated that the outcome can be accepted.

Validation 2:

Comparison between the the histograms of stones per sample for the data and simulation (Figure 5.2) is favourable. In the figures are the conditioning data (in blue) and in red, the outcome of the simulation. Figure 5.2 (right), shows the comparison in more detail close to the origin.

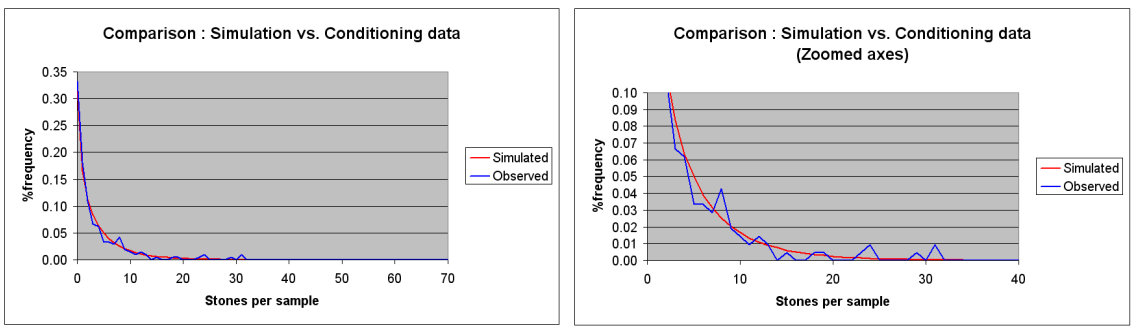


Figure 5.2: Comparing the conditioning data with the simulation outcome

Validation 3:

The simulation results were regularised into blocks of 50x50m and compared to the 50x50m production data. The table below shows the comparative statistics.

Table 2: Simulation results on 50x50m block scale, compared to production data

Statistic	Production data (50x50m blocks)	Simulation (regularised 50x50m blocks)
Mean	912	844
Variance	392 201	432 275

The data from the simulation blocks covering the same extent as the those blocks mined as production data, compares favourably. The slight difference between the sample grade in table 1 and that of table 2 is due to the fact that the mining was focussed on a selected region with a higher average grade than that of all the sample data.

Validation 4:

The most important validation is to check the variography behaviour of data based on samples less than 10m^2. The deconvolution of the variogram was not satisfactory and this study requires an analysis of samples smaller than 10m^2 .

The sample sizes considered for the smaller sampling tool were: 0.72m^2, 2.16m^2, 3.14m^2, 5m^2, 7m^2, 10m^2, 20m^2 and 50m^2. An analysis was done using each of the sizes and the results interpreted.

The simulation was extensively sampled (40 000 samples over a 2x2km area) for each sample size and variograms modelled for grade. The variogram's nugget effect should change from one support to another by the factor v_{small}/V_{large}. The theoretical behaviour of the variograms and variograms based on sample observations from the simulation, were compared. A summary of nugget effect to sill ratio is shown in the table below.

Theoretical versus modelled ratios of Nugget Effect to Sill

Sample size	Ratios determined from theoretical scaling (v/V)	Ratios observed from sampling the simulation
0.72m^2	0.82	0.82
2.16m^2	0.63	0.58
5m^2	0.49	0.50
7m^2	0.43	0.43

In all cases the variogram ranges were similar and modelled as 96m. The small differences between the simulation and theoretical expectation indicates that the theoretical expectation and observations made from the samples smaller than 10m^2 in size are in agreement with one another. Although the deconvolution of the statistics for use in the simulation could not be completed successfully, the variograms at support sizes smaller than 10m^2 had sill to nugget effect ratios according to theoretical expectation.

For all practical purposes, the spatial behaviour of the simulation at a scale less than 10m^2 was found acceptable.

5.3.6 Creating a low grade simulation

De Beers Marine aims to mine progressively lower grades and increase the mining rate to make more of the resource available and, as a result, more financially viable for exploitation.

Only carrying out a sample optimisation study on a relatively high grade area would introduce bias. For this reason an attempt was made to create and analyse a low grade simulation to represent future mining and sampling in such areas.

The lack of data from low grade areas required an assumption to be made that similar depositional processes were present both in the high and low grade areas. It was shown in studies that the process of removing stones (decimation) can be performed without compromising the validity of the Cox process, mimicking the determination of parameters for a Cox process at a lower grade, in order to scale high grade simulations to a lower grade for use in spatial studies.

Following this reasoning, the conditional Cox simulation was seeded with:
a) The original grade and
b) Grade scaled with a reduction from 0.33 stns/m^2 to 0.15 stns/m^2

The low grade simulation was created as a subset of the high grade simulation by randomly eliminating stones in the seeding process using a factor of 0.44 to reduce the grade.

The two realisations of the simulations are shown in Figure 5.3 where each dot represents a diamond occurrence on the seabed. A visual check of the conditioning data, simulation results and seeding process showed that the simulations were acceptable.

The simulation realisations were accepted as a reasonable representation of the spatial location of individual diamonds on the seabed and were used for the optimisation study.

5.3.7 Sampling

For this study only a single simulation was created for use in the analysis. A comparison to test the effect of increasing sample support is required and for this reason the samples were placed at fixed locations. This eliminates the variability associated with sample location variation. The fixed sample locations with changing sample size is demonstrated in the figure below.

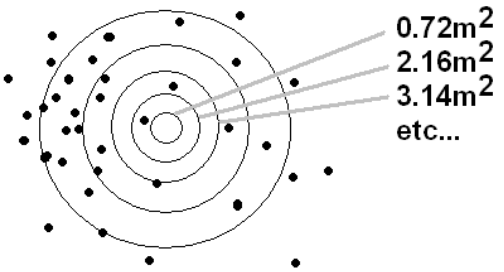


Figure 5.4: Increasing the sample size, located at fixed sample locations, at the midpoints of 50x50m blocks

This enables block estimates and sample representivity to be directly correlated and compared with one another.

Investigating the representivity of samples has the objective of testing for accuracy and unbiasedness of results. The sampling and estimation method followed does not take into consideration sample contamination, sample location errors, loss of stones in the recovery process, over penetration or any other errors that may be introduced.

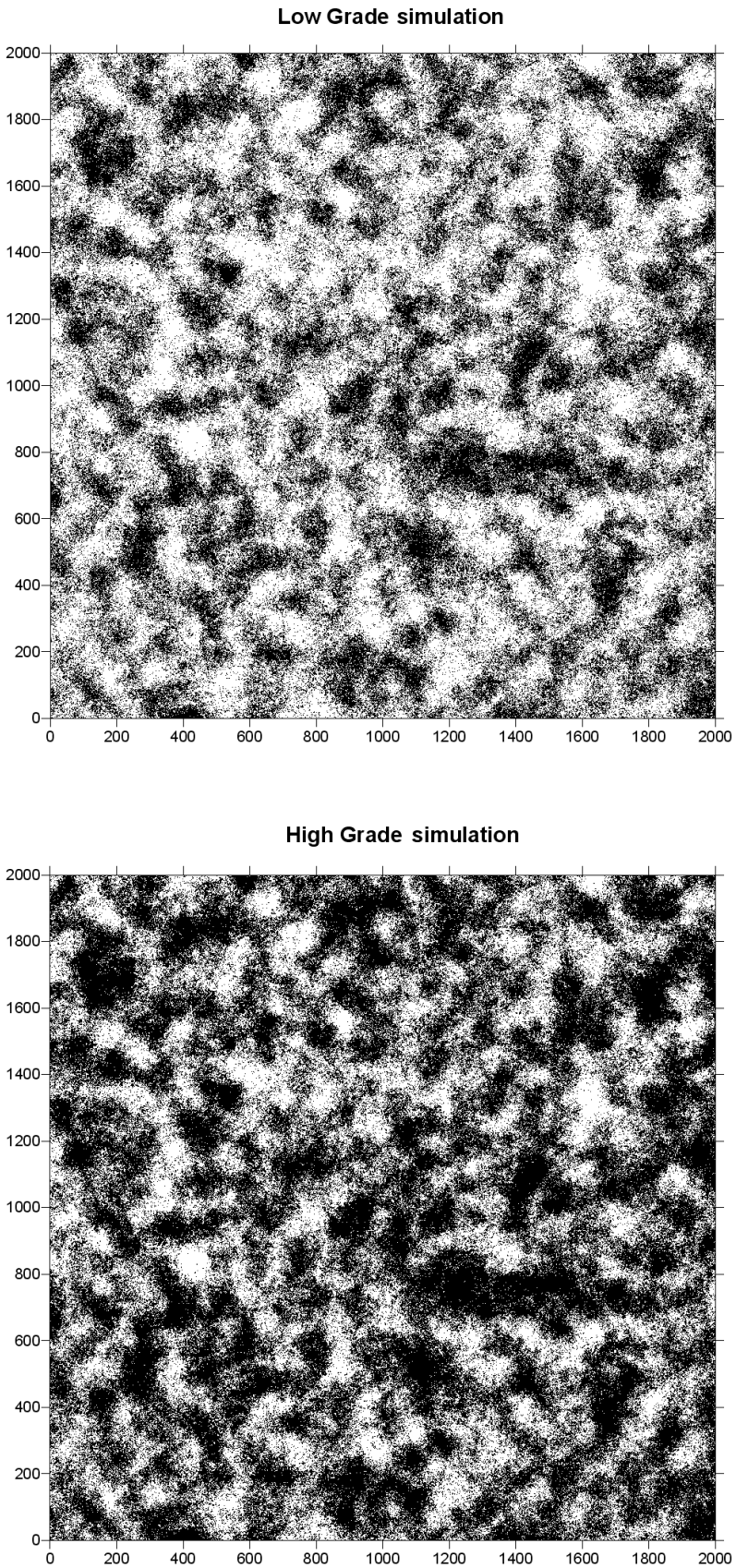


Figure 5.3: High and low grade conditional Cox simulations of a marine diamond placer deposit

5.3.8 Analysis using the simulations: Variograms, kriging results and interpretation

For the graphs and tables that follow, two sets of results are shown: The first is obtained from analysing the high grade simulation and the second from an identical study on the low grade simulation.

Calculating variograms:

Each of the sample support sizes was used and sampled on a regular 50x50m grid. An omni-directional variogram was modelled using spherical structures for each of the outcomes.

The modelled variograms are plotted in Figure 5.5. For some plots, it was impossible to model a variogram and when compared to the expected theoretical variogram the structure was poorly represented. Such sample support sizes are considered too small to reflect the true underlying spatial structure. For the low grade simulation, variograms started displaying spatial structure and were modelled from sample sizes of $7m^2$ and upwards. For the high grade simulation sample sizes of $5m^2$ and larger could adequately represent the underlying variogram structure.

Those variograms that do not model the underlying spatial structure sufficiently accurate were greyed out in Figure 5.5.

Estimating 50x50m blocks: Kriging results

Using the modelled variograms for each sample size and based on comprehensive datasets (40 000 samples, spaced at 10x10m intervals), the slope of regression for a theoretical kriging exercise based on 50x50m spaced data was calculated. The results for a 3x3 kriging neighbourhood are shown in the table below:

Table: Sample size and theoretical slope of regression for kriging

Sample size (m^2)	Low grade: Slope of regression	High grade: Slope of regression
0.72	0.23	0.38
2.16	0.47	0.66
3.14	0.57	0.73
5	0.66	0.80
7	0.75	0.85
10	0.81	0.89
20	0.89	0.93
50	0.96	0.97

In ideal conditions, with an abundance of data and a variogram structure with low nugget effect ratio, a slope of regression of approximately 0.9 or higher is preferable for kriging. It is evident from the table above that this occurs only once samples are $10m^2$ or larger.

Kriging was performed on each of the two simulations sampled on a 50x50m grid and using a 3x3 kriging plan. The average slope of regression was calculated from individual block results.

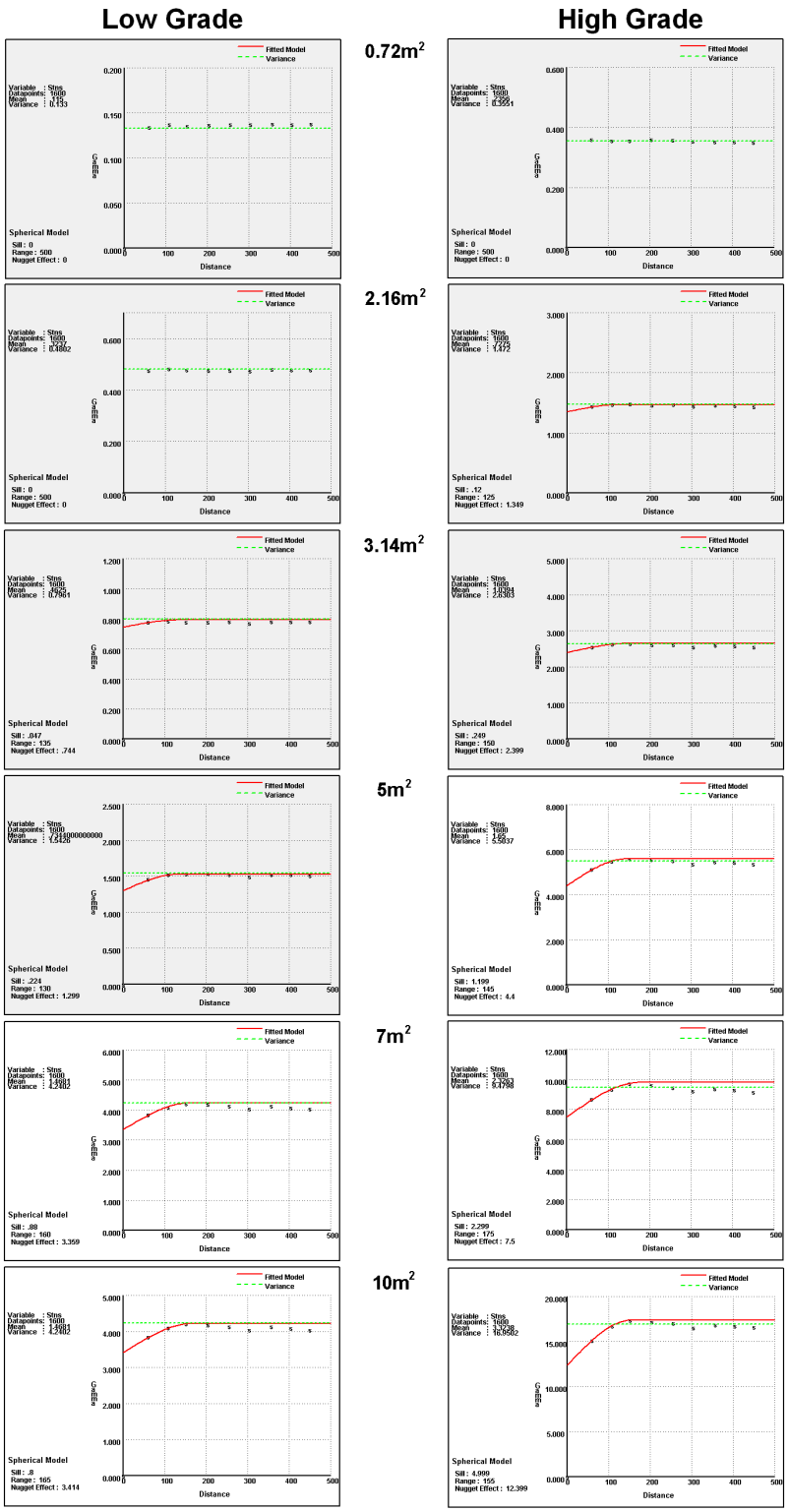


Figure 5.5: Variograms for data on 50x50m sample spacing for low grade (left) and high grade (right) simulations

A 5x5 kriging plan was also considered and a similar theoretical slope of regression calculation as for the 3x3 plan was repeated. The results of the 3x3, 5x5 kriging plans and observed slope of regression are presented relative to the 0.9 slope in Figure 5.6.

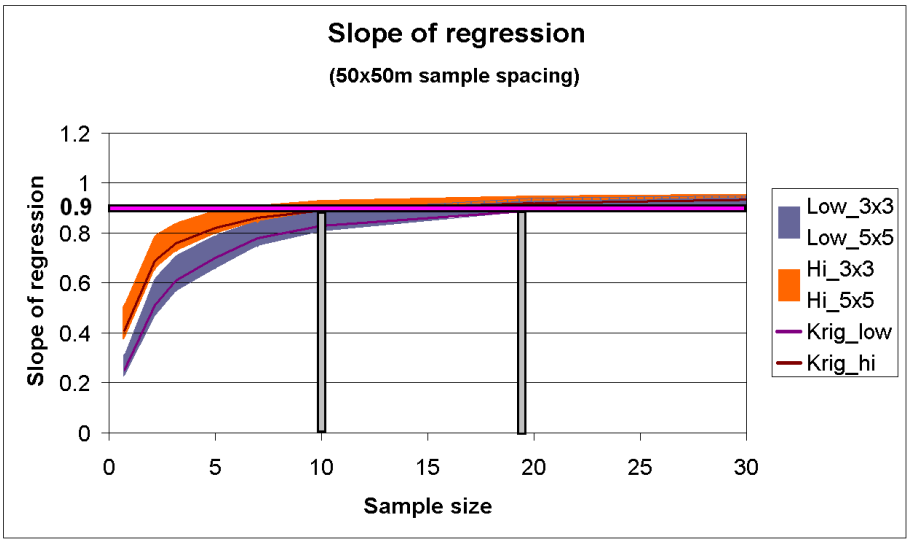


Figure 5.6: Slope of regression from kriging with 3x3 and 5x5 plans

The slope of regression calculated for the high grade simulation approaches 0.9 at a sample size around 10m^2 but for the low grade simulation this is around 20m^2 . If 0.9 is a benchmark slope of regression to achieve, a minimum of at least 10m^2 or larger sample would be required.

Correlation coefficients: 50x50m blocks

Correlation plots between "actual" and estimated were prepared in Figure 5.8. A summary of the correlation coefficients is shown in Figure 5.7:

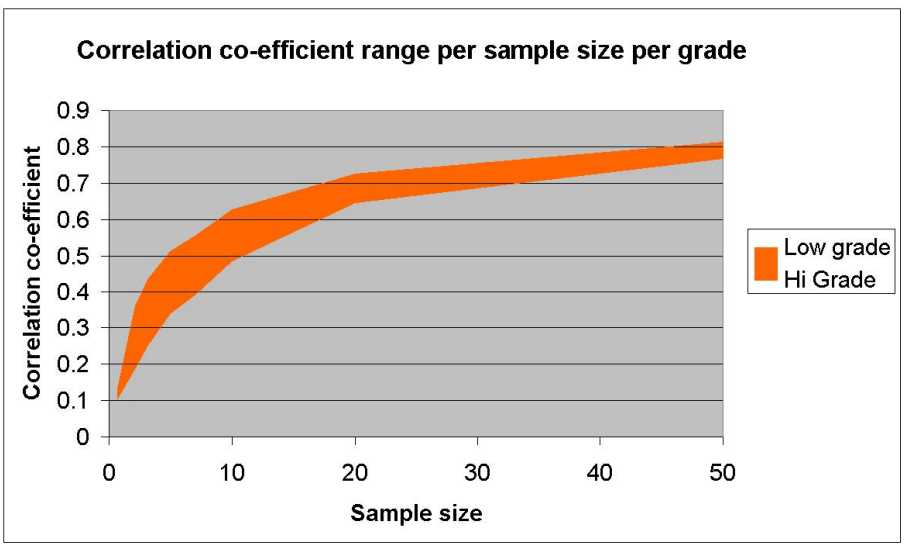


Figure 5.7: Correlation per sample size for low and high grade simulations

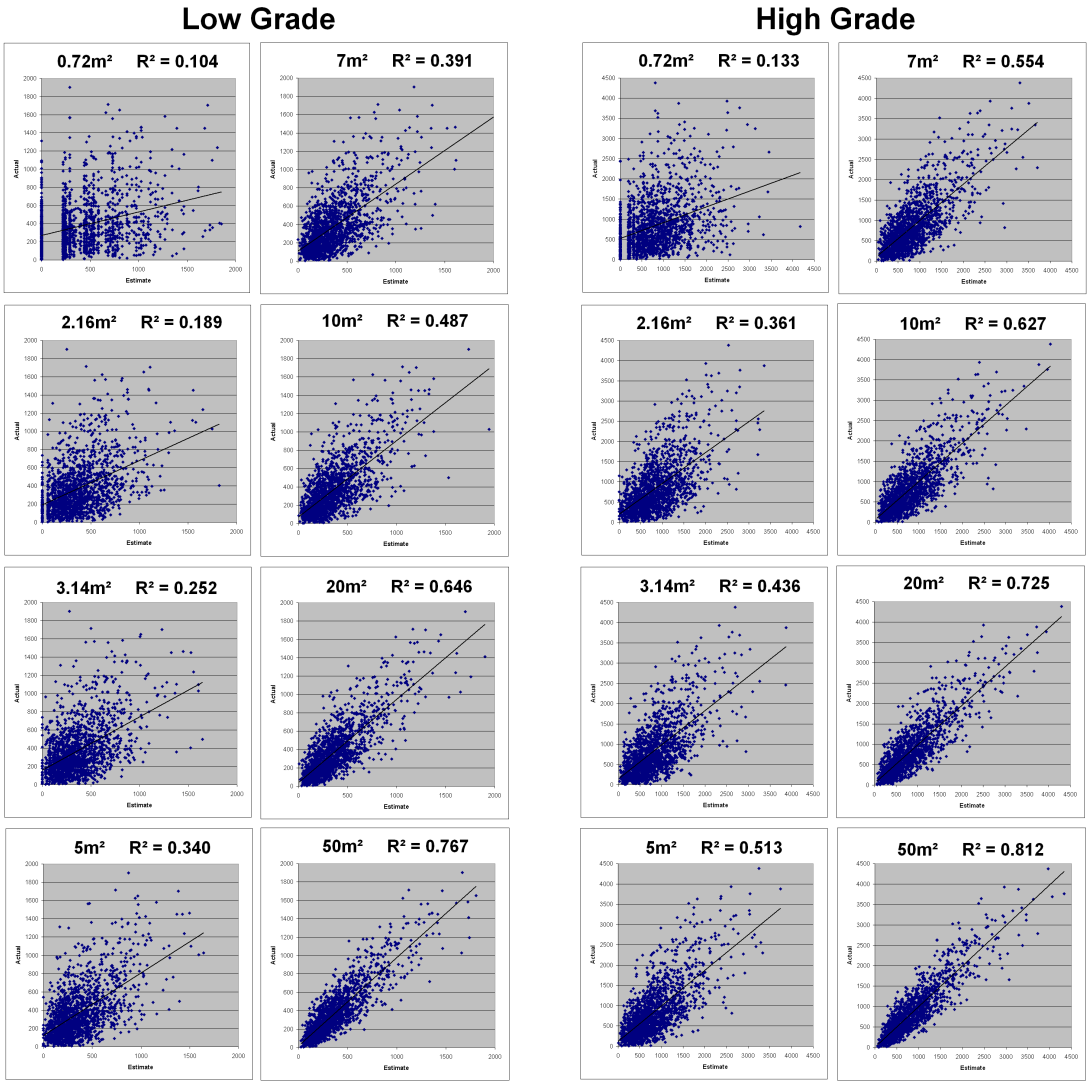


Figure 5.8: Correlations plots for 50x50m blocks per sample size. Low grade (left) and high grade (right) simulations. *Axis legends: Y-axis = "Actual" and X-axis = Estimate*

Figure 5.7 shows that correlation increases fastest up to a sample size of around 20m^2 and after that the increase is marginal with increasing sample support size.

The correlation clouds in Figure 5.8 show that it is only at larger sample support sizes (>20m^2 for the low grade simulation and 7-10m^2 for the high grade), that the correlation clouds become cigar shaped and a more meaningful correlation is observed.

Estimating panels of 200x200m: Correlation coefficient

What is important as a measurement for the accuracy of estimates during the mining process is not the results of individual 50x50m blocks (although these are also monitored), but the accuracy of mining panels of 200x200m. The ship anchors on a spread and a panel of 200x200m is mined. The kriged results across the 2x2km area were gridded into panels of 200x200m and compared to the simulated "actual" and the results are shown as regression plots in Figure 5.9.

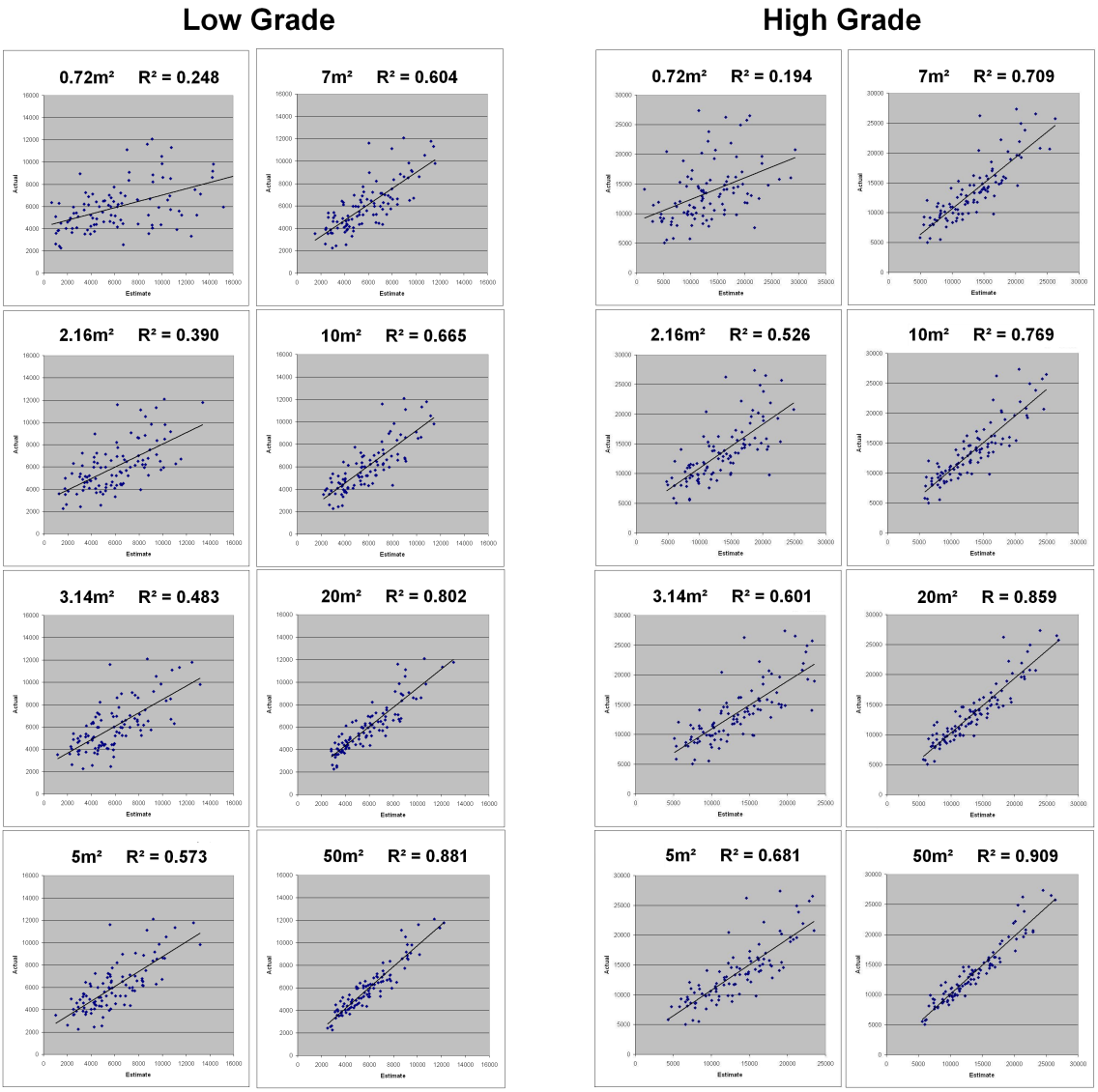


Figure 5.9: Correlation plots for 200x200m panels by sample size, for low grade (left) and high grade (right) simulations

On a panel scale, the correlation becomes more meaningful to interpret than for the individual 50x50m blocks as it has more relevance to the scale on which mining is planned. A summary of the correlation coefficients is shown in Figure 5.10.

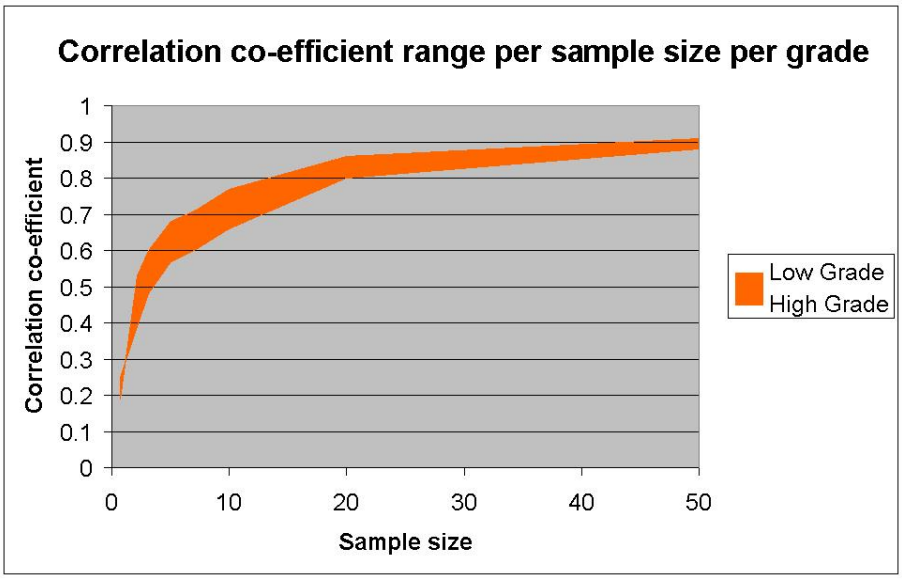


Figure 5.10: Correlation coefficients on 200x200m panel scale, by sample size for low and high grade simulations

The graph in Figure 5.10 shows a rapid increase in correlation (up to $r^2 = 0.6$ to 0.7) for sample sizes up to $5m^2$, then become gradually more for sample sizes up to $20m^2$ and even more gradual for sample sizes above $20m^2$.

Accuracy of 200x200m panel results

Figure 5.11 was prepared by averaging the kriged results from 50x50m blocks into 200x200m panels. These results are expressed as a percentage of accuracy relative to the actual panel grade calculated from the simulation. For this study, 15% and 20% were selected as acceptable deviation from the "true" values as the criteria against which to measure the accuracy of results. Figure 5.11 plot the percentage of panels falling within the 15% and 20% limits of accuracy relative to the "true" grade. According to the risk profile that management is willing to take, an acceptable sample size can be derived.

The graph indicates that for samples smaller than $10m^2$, the risk of having results outside a 15% deviation from the true grade (on panel scale), increases steadily and, even at a $20m^2$ sample size, the mining results show many of the panel results outside acceptable limits.

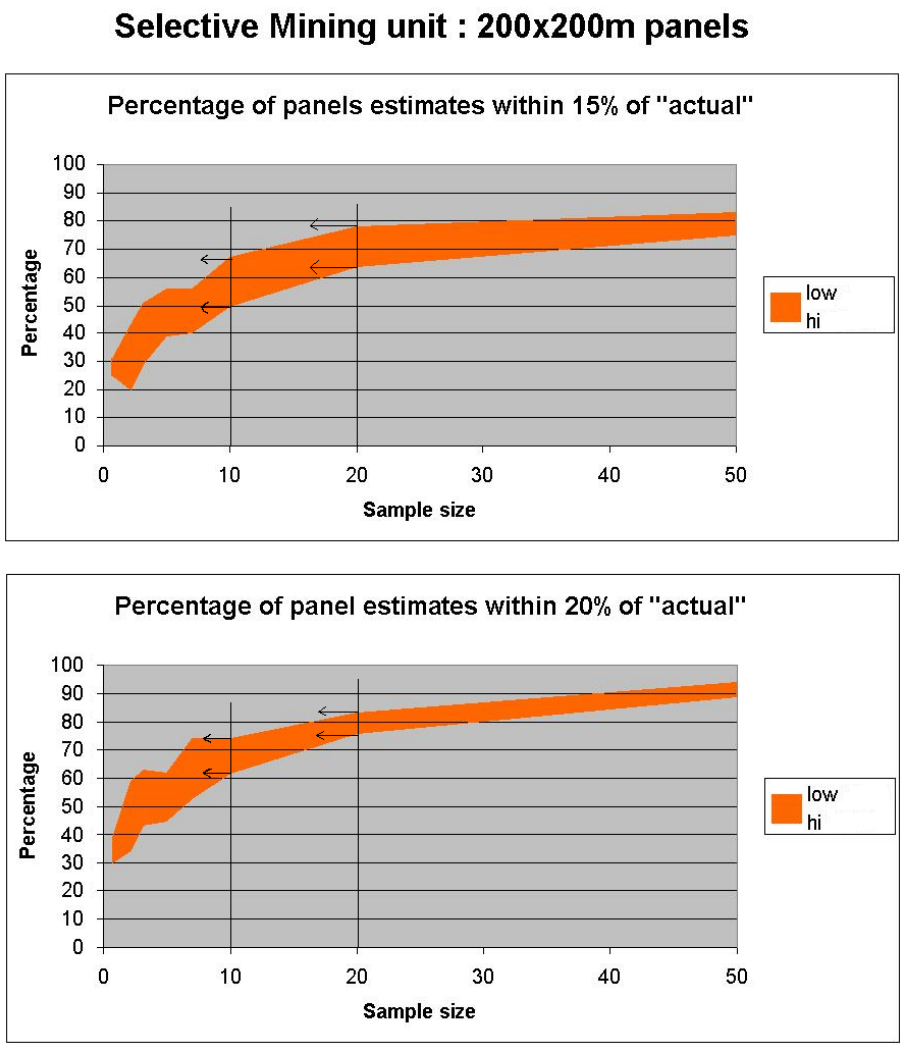


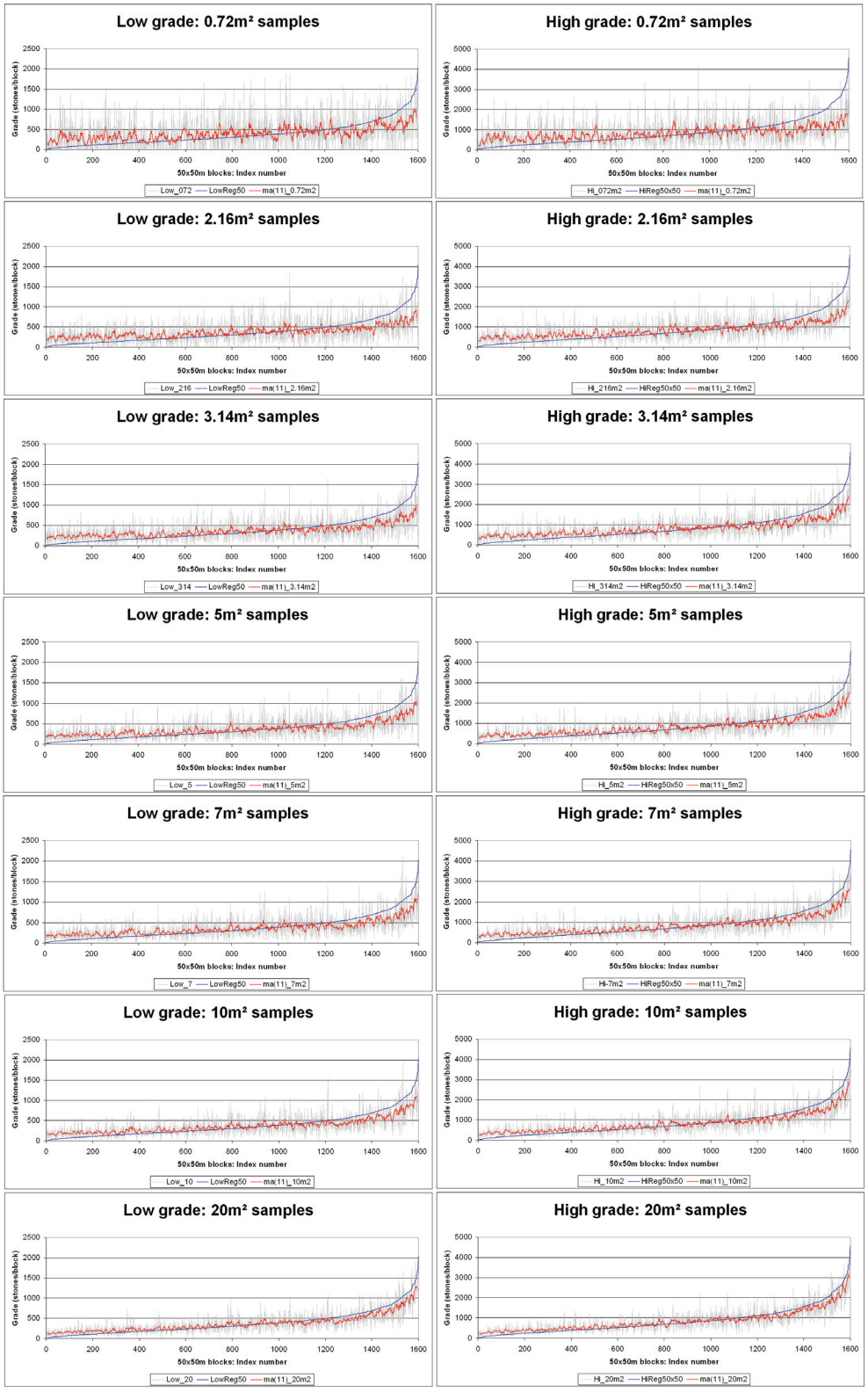
Figure 5.11: Summary showing estimation accuracy for 200x200m panels

This variation around the true panel grade could be problematic within the framework of resource classification schemes and project assurance for funding. Fifteen percent accuracy is often quoted as the percentage against which the accuracy of estimates is benchmarked. Analysing the simulation results on block and panel support, a sample size of less than $10m^2$ is not recommended.

Accuracy of 50x50m block results

Observations from historical data show that small samples tend to underestimate the high grade blocks of a deposit according to Duggan [20]. The conditional bias also causes the low grade blocks to be estimated too high. The same observation was made in this study as shown in the graphs in Figure 5.12. The estimates are plotted on the Y axis against an index number on the X-axis, based on the 50x50m block results sorted from the lowest to highest grade.

The estimates data is smoothed with an 11 point unweighted moving average to highlight the trend of estimates against "actual". This graph shows that at a sample size of approximately



e

Figure 5.12: Comparison: Estimates and "actual" for 50x50m blocks, low and high grade simulations using sample sizes: $0.72m^2$, $2.16m^2$, $3.14m^2$, $5m^2$, $7m^2$, $10m^2$ and $20m^2$

7m^2 (for high grade) and 10m^2 (for low grade) the estimation curve (in red), coincides with the actual (in blue). This suggests an acceptable estimation accuracy of the higher grade blocks.

Finding these higher grade blocks through accurate estimation is essential for the selection process and inclusion as part of the mine plan.

The interpretation of estimation accuracy for 50x50m blocks and 200x200m panels show that 10m^2 is the smallest sample support size (assuming a single sample on a 50x50m grid) which will produce an acceptable result for doing grade estimates.

5.4 Summary and conclusions

Research to test the possibility of implementing a sample tool less than 10m^2 for use in the placer environment was done. For this purpose a conditional Cox simulation was used.

With only 10m^2 sample support data available, the input statistics had to be deconvoluted to point support. The deconvolution to a sample support of less than 10m^2 could not be done satisfactorily and a simulation was created on a sample support of 10m^2 . Various validations were done on the simulation to test if this approximation would be acceptable for a study on sample supports less than 10m^2 . The tests showed that the simulation was acceptable. A second, low grade simulation was created to establish the effect of sample support in lower grade areas.

A single realisation of the Cox simulation was analysed and samples with different supports were sampled on a grid centred at the midpoint of 50x50m blocks. These blocks were estimated using kriging and averaged into panels of 200x200, the same size as mining panels.

Analysing the results on 50x50m block scale, looking at variogram reproduction, correlation coefficients, theoretical slope of regression, and variation to "actual", in addition to the variation on 200x200m panel scale relative to a benchmark of 15% to the actual block grade, has confirmed that a 10m^2 sample is considered the optimal sized sample to use in the Atlantic I placer deposit.

The conditional Cox as described in the appendix was used as the simulation model with which optimisation studies can be successfully created in marine environments.

Chapter 6

Sampling and estimation of diamond grade in tailings resources

6.1 Introduction

The diamond recovery process is not perfect for various reasons and some diamonds are discarded along with the tailings material. The recovery process is furthermore designed to treat material at specified economic cut-off sizes to avoid the treatment of large volumes of material containing only small diamonds of nominal value. Over time, large resources of mined material that still contain diamonds have been formed. Most diamond mine tailings dumps can be assigned a grade and are therefore referred to as a tailings mineral resource (TMR). Typically, historical tailings deposits may become viable because of changes in economic demand. This is also made possible by the development of improved metallurgical recovery processes.

Evaluating a tailing mineral resource requires a sufficiently accurate estimation of the diamond content and associated revenue. An estimation method is proposed for sampling these resources in order to estimate local block grades in diamond-bearing dumps before exploitation.

6.2 Why are diamonds present in the TMR's?

Generally, diamonds occur in TMR's as free (or partially free) particles that have been either completely liberated or partially liberated from the host rock or occur as a locked-up diamond in conglomerated material or material that has not been adequately crushed.

Some reasons for the occurrence of diamonds in dumps are:

Metallurgical process The density of diamond is 3.52 g/cm^3 and the metallurgical process utilises this character by creating a medium from which the material with a higher density can be extracted. When a piece of ore with a density that is less than that of a diamond and of sufficient size is still attached to, or encapsulating the diamond, it can cause the

particle to float of as waste in the Dense Media Separation (DMS) process. The DMS process referred to is a combination of the density of the medium, as well as the vortex created by the cyclone. Schematically:

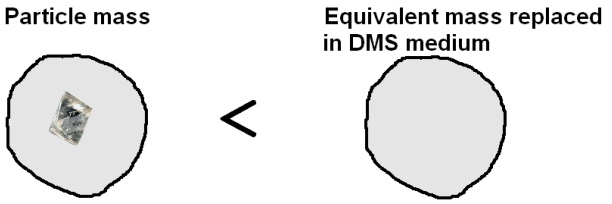


Figure 6.1: A diamond locked up and floating off as waste

If the combined weight of a particle and diamond is more than the mass of the volume it replaces in the medium of the DMS, the particle will sink and be recovered, else it will float off as waste. From work done by Ferreira [24] and Lantuéjoul [45] it can be represented as follows:

K_d as Kimberlite density,	K_v as Kimberlite volume	
D_d as Diamond density,	D_v as Diamond volume	and
M_d as Medium density,	M_v as Medium volume	

For a particle to float in the DMS:

$$(K_d \times K_v) + (D_d \times D_v) \leq M_d \times M_v$$

With $K_v = M_v - D_v$ if follows that:

$$D_v \leq M_v \frac{M_d - K_d}{D_d - K_d}$$

By substituting appropriate parameters for the metallurgical process, a value can be determined for the largest diamond volume [D_v = volume $\equiv$ weight $\equiv$ size] that can be floated out of the DMS process, giving an indication of the potential diamond lock-up and subsequent loss of diamonds to the TMR's.

The cut-point of the DMS is not absolute and particles will float off according their densities described by the Tromp curve. The particles will be floated off according to a statistical probability distribution and any deviation of the curve or statistical fluctuation of the parameters controlling the DMS process will cause diamonds to float off as waste.

Feed rate and DMS efficiency The ratio of reduction of material through the DMS process is around 100:1, and if at any stage the density of the cyclone deviates and increases from the ideal cut point for example, material surging through the process, overloading of the cyclones can occur and free diamonds can float out as waste. Also, if the shape of a diamond is highly aspherical, for example a flat diamond (macle), it has a good probability of floating out as waste.

Financial benefit of diamonds below a minimum cut-off size Material is crushed or milled in stages down to $6mm$ (occasionally this may be $9mm$ or $12mm$, depending on the material type, diamond size and results from tests done to relate crusher size gaps to the percentage of diamond recovery). Typically, material generated in the size fractions less than $1.5mm$, is not treated and small diamonds in this size fraction are discarded. In a mass-balance of the metallurgical process, this can be around 50-60% of the total volume of material (personal communication : Stephen Coward: metallurgist). The diamonds in this small size fraction cannot be recovered economically due to their market value not covering the associated treatment cost.

X-ray and grease belt efficiency X-ray equipment is calibrated according to the expected luminescence of the diamonds, but deviating from the calibration parameters can cause a diamond with very low luminescence to be lost. Some diamonds do not luminesce and in mines where grease tables are not used (making use of the hydrophobic properties of diamonds instead of luminescence), diamonds would be discarded. If a mine utilises grease belt recovery only (common in historical operations), and diamonds are coated with a microscopic non-diamondiferous film which make the diamond lose it's hydrophobic characteristics, these diamonds get discarded to tailings.

Washing pans This method of diamond recovery was and is still used and have a medium to high recovery efficiency according to BRC Diamond Core Ltd. [9]. Highly skilled persons with lots of experience can make this process very efficient, but generally, the use of this technology loses diamonds in the recovery process. This makes dumps that were created before DMS technology was introduced (with generally higher recovery efficiency), a potentially viable resource of diamonds.

Lock-up Diamonds completely encapsulated by host rock are not exposed to X-rays and will not fluoresce. Such stones cannot be recovered using X-rays nor by grease belts. After years of natural weathering on the TMR these diamonds may occur as liberated diamonds.

Liberation Liberation studies are undertaken on the size of the final crushed product that will release the optimal number of diamonds. The following extract from a technical report illustrates expected recoveries in a placer mineral resource[1]:*".... it is expected that 87% of the stones would be recovered at the first stage after only screening off the finer* $-1.4mm$ *fraction. About 5% of the diamonds would be recovered after crushing to* $-12mm$ *and a further 4% after crushing to* $-9mm$*"*. Diamonds not liberated by the comminution process are therefore discarded onto the TMR.

Understanding the reason(s) for the diamonds being on the dump and time periods over which metallurgical processes affected the deposition, will help in the design of a more efficient plant to treat the material. Although there are diamonds present on the dumps, the recovery processes are continuously monitored to minimise this from happening and over time, the grade of dumps generally decline.

[1] *Evaluation of bulk sampling requirements for the #2 plant tailings dump* by C F Prins, S P Duggan, M Loubser & G Viviers, De Beers Internal Report, September 2003

Figure 6.2: Low grade TMR's derived from mining placer deposits. Seen from a height of 1.3km

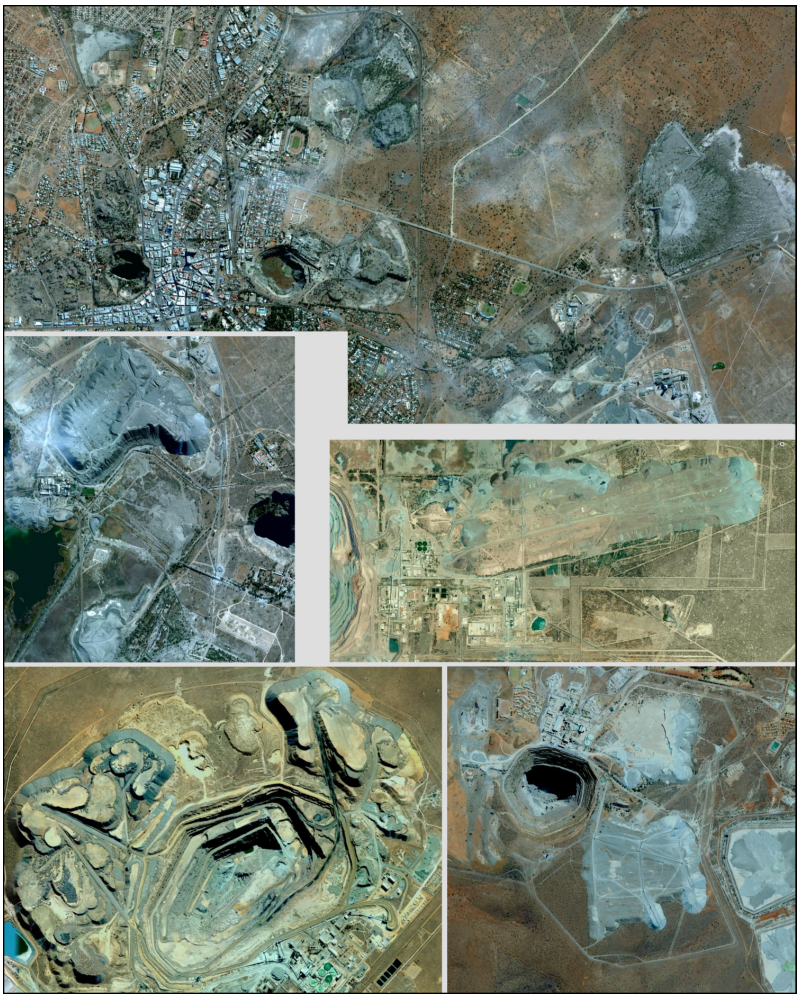


Figure 6.3: Shapes of kimberlite tailings TMR's. Seen from a height of 5km

6.3 Types of tailings resources, with examples

Different types of tailings resources could be considered for sampling. These deposits consist of material that has been subjected to a recovery process whilst stockpiles comprise material that has been stored for treatment at a later stage. Below are descriptions of different types of TMR's and stockpiles:

Tailings resources (the focus of research for this chapter) These resources are created from material that has been processed for diamonds and the size of the material is between a maximum crushed size (for example $+9mm$) and a minimum size (for example $+1.5mm$).

Slimes dumps All material smaller than the minimum particle size (around $-1.5mm$), considered to be of no economic value.

Metallurgical Stockpiles These stockpiles are created with material that cannot be treated. For example at alluvial mines that do not have the ability to crush, where cemented material of significant proportions are found as layers within unconsolidated material (for example, historical stockpiles at Elizabeth Bay mine near Lüderitz: Namibia).

Plant Stockpiles These stockpiles are created during mining operations where the mining is faster than the recovery process and/or deliberately stockpiled material for periods when rain disrupts activities in open-pit operations (Example: Debswana mines in Botswana).

Low grade stockpiles Stockpiles of material that were considered uneconomic to mine (under different historical financial conditions). These stockpiles are now reassessed through sampling and estimation to determine if this is still the case.

Overburden stockpiles These are more specific to placer deposits, where very low grade, typically sandy deposits with some mineralisation overlaying the higher grade gravels are stripped of, screened and placed on a stockpile. These are flagged as potential resources for future extraction.

Red area dumps These are tailings material stored inside the red area (high security area) after a process of hand sorting of the material in the larger size fractions of, for example: $+32mm$ to $-60mm$ material. X-ray tailings also fall in this category.

Stockpiles form part of the mining process and is seldomly sampled, whereas TMR's, the focus of research of this chapter, require an approach for the sampling and estimation. Examples of TMR's are shown in Figures 6.2 and 6.3. Tailings deposits vary from a few hundred tons to many millions of tons of material to form minable resources. The grade and latest financial assessment will determine if it is economic to mine.

As examples and to provide some insight into the potential grade of tailings resources, the following two examples are presented.

Example 1: Alluvial tailings dump

The first example is of composite samples taken over a two year period from the #4 plant dump near the town Oranjemund in Namibia as documented by Loubser[48], with the purpose of estimating a global grade. Using this information, a decision had to be made whether

the sampling described the global grade accurately enough with the current information, or to design an additional sample campaign to determine the grade more accurately, keeping in mind that these types of dumps are sometimes so marginal that the cost of another comprehensive sample campaign may cause the tailings resource to become unprofitable to mine.

The sample results (location and size of sub-samples) from a programme conducted in 1996/97 are shown by Loubser [47]:

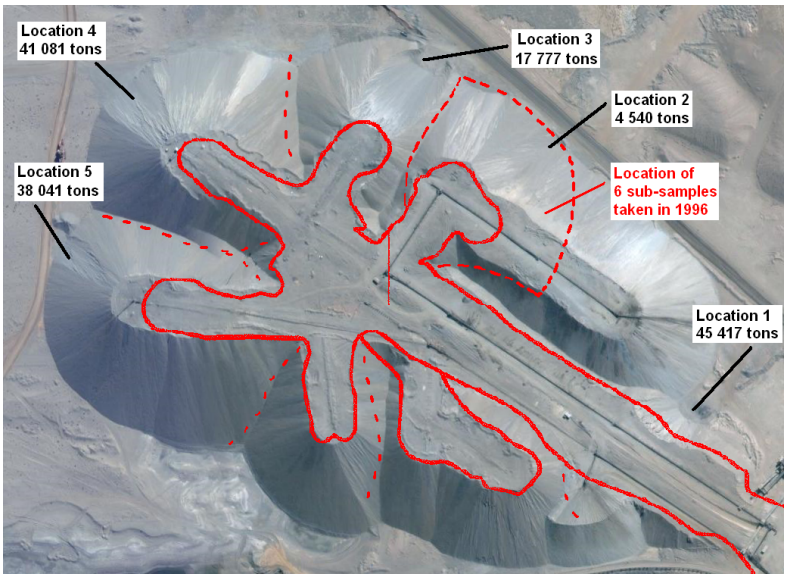


Figure 6.4: Location and tonnages of samples relative to the 1996 dump outline

A summary of the sample results was taken from Loubser[48]:

Table 6.1: Tons, carats & stones recovered for the 1996 and 1997 sample programmes

Date tipped	Tons	Carats	Stones	Grade (cpht)	Size (cts/stone)
06/06/96	500	5.05	11	1.01	0.459
05/07/96	500	0.45	1	0.09	0.450
16/07/96	500	0.43	1	0.08	0.430
30/08/96	575	0.58	2	0.10	0.290
18/12/96	410	8.40	27	2.04	0.311
20/12/96	450	5.66	15	1.25	0.377
30/10/97	(*)146 856	386.08	609	0.26	0.634
TOTALS	149 791	406.65	666	0.27	0.610

(*)This sample is a composite of the 5 sub-samples (locations 1 to 5) shown in Figure 6.4

The sample results shown in table 6.1 was used to estimate the grade. The "samples" have low levels of integrity because they were treated through a production plant facility. The average grade was marginally below the economic "break-even" grade however, a decision was required to mine the TMR or, as determined at that time, to postpone mining to improve the confidence by further sampling.

A (geo)-statistical sampling methodology is required as a benchmark against which the current data can be measured.

Example 2: Kimberley Central Treatment Plant (CTP) tailings dump

The tailings resource in Kimberley was sampled in 1976 and 1977 to determine the grade of the dump. Below is an extract from Kilham [36] of historical data collected as 12m long samples drilled from the top surface of the dump.

Table 6.2: CTP historical sample data : Williams Auger data

Sample #	From	To	Carats	Stones	Sample #	From	To	Carats	Stones
44A	0	12	0.7	17	continued...				
44B	12	24	1.1	21	75A	0	12	0.7	20
44C	24	36	1.1	18	75B	12	24	0.6	18
45A	0	12	1.4	19	75C	24	36	0.8	14
45B	12	24	1.6	12	76A	0	12	3.3	29
46A	0	12	0.5	14	76B	12	24	1.2	22
47A	0	12	0.7	15	76C	24	36	1.7	25
47B	12	24	0.4	14	77A	0	12	1.0	15
47C	24	36	0.6	15	77B	12	24	0.7	17
48A	0	12	0.4	14	77C	24	36	1.4	27
49	0	12	2.1	34	78A	0	12	1.4	25
49A	0	12	0.6	14	78B	12	24	0.8	17
				etc...	78C	24	36	1.2	22

With this data available, research was undertaken to establish an optimal sample campaign taking cognisance of the fact that the data is of very low confidence. An extract below was taken from an internal De Beers report to highlight some problems with the data [2].

... The 1976/1977 Williams Auger data is considered to be of poor integrity and unsuitable for mineral resource estimation for the following reasons:

- *The samples were treated in a bulk sampling plant attached to the NTP, and final recovery took place in the production sort house (John Kilham, pers. comm.). Final diamond recovery in a production facility is considered to be poor practice, and can easily lead to contamination of the samples.*
- *The sample grades are very consistent with depth – this high degree of consistency is viewed with some suspicion, in that downhole contamination, and consequently homogenisation of the grades, may have taken place. This is supported by reports that the Williams holes often suffered collapse where drilling conditions were adverse (e.g. Boshof Road; Clement, 1977).*
- *The Williams holes only penetrated to a depth of 36m, and therefore have not sampled the bottom half of the TMR.*
- *Details concerning the exact treatment parameters, e.g. recrush parameters, are not clear. This makes it difficult to align the 1976/1977 Williams results with the BG36 campaign.*
- *The Williams Auger hole diameter was 1m, and 12m lift samples were collected; this is considered to be sub-optimal with respect to individual sample support size, and their inclusion would introduce a mixed support problem.*
- *The Williams Auger hole area of influence is covered by the BG36 campaign.*

The Williams Auger results have therefore been disregarded for the purposes of this mineral resource estimation study...

[2] *A Mineral Resource Estimate for the Current Tailing Mineral Resource, Chronofacies 1, Kimberley Mines*, by MG Millad, 6th July 2006

The extract from the report highlights the problems with the data and suggests it not being used for a resource estimate. An alternative approach, not utilising this data, will have to be found to optimise the sampling.

This CTP dump, with historical information, is the topic for the case study of this research.

6.4 The metallurgical challenge to recover diamonds

If the expected grade of diamond bearing material is high, a relatively small sample is required to determine the presence/absence of diamonds.

In the case of very low grade dump material (for example tailings resources derived from placers on the West Coast), the "extremeness" in the practicality of the measurement of grade can be demonstrated in the following way:

For an average grade of 0.3cpht with an average stone size of 0.5 cts/stn (which was a marginal economically minable dump at the time), a single $5m^3$ truck filled with gravel will weigh about 8 tons. If twenty of these trucks are parked alongside each other, there will be only one stone in this volume of material at the assumed grade. If this single stone is recovered by the process, the estimate of the mineral resource will be close to an economical break-even point and two stones will result in a forecasted profit.

Pictorially:

Figure 6.5: A total of 20 trucks, each filled with 8 tons of gravel, containing a single stone

To emphasize the metallurgical recovery efficiency, the recovered product is shown!

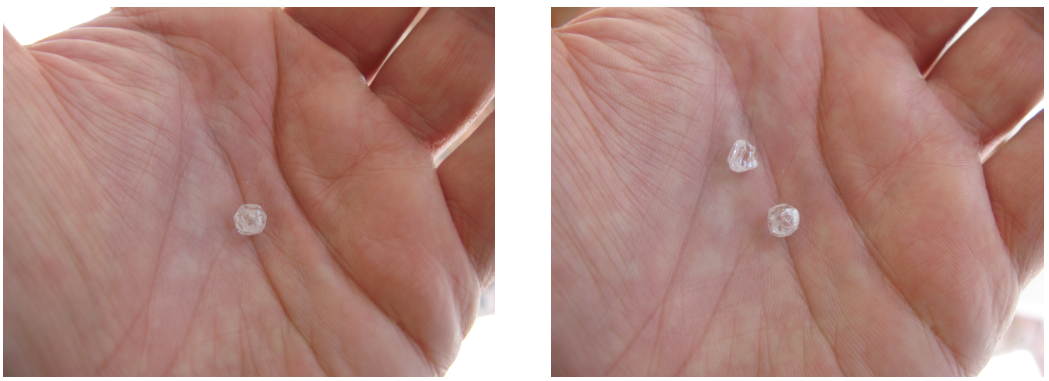


Figure 6.6: Break-even cut-off grade (left) and above cut-off grade (right)

This demonstration can be considered an extreme example because it is a very low grade

alluvial tailings mineral resource but serves to highlight the challenges related to sampling and evaluating TMR's.

It specifically demonstrates how easy the loss of one or more stones in the metallurgical process will have a significant impact on the grade estimates.

6.5 Background information

In 1999 De Beers initiated discussions to establish an appropriate sample size and configuration for the evaluation of tailings resources. In 2004 the focus was shifted to improving the confidence associated with the estimates at the Kimberley Mines and specifically the Kimberley CTP (Central Treatment Plant) tailing mineral resource (TMR). In addition, De Beers needed to find a method to estimate local block grades within the dump. The shape and size of the TMR is shown in Figure 6.7 below. The CTP TMR has an area of 1 200 000 m^2 and a tonnage of 120 million tons.

Figure 6.7: The Kimberley CTP dump

A formalised approach to the sampling of TMR's was established, where the method accommodates low grades and complex genesis. The sampling program was required to deliver an optimum sample configuration whilst taking cognisance of practicality, budget, time constraints and an appropriate degree of accuracy.

Although the method was specifically developed for the CTP, the method needed to be of a general nature for application to all TMR's.

The geostatistical principles of spatial sampling and estimation of in-situ material are not

applicable to TMR's because the structure are non-stationary and the shape too difficult to address with geostatistical methods. The problem required the identification and solution of specific issues that include:

- A review of the depositional processes of TMR's
- Identification of the main reasons for diamond occurrence
- Establishing an appropriate sample size
- Identification of homogeneous zones within the TMR
- Establishing an optimal sample space configuration, and
- Creation of a methodology to estimate local block grades

6.6 Determining the sample spacing

Researching a method to determine optimal sample spacing cannot use horizontal or vertical spatial relationships between samples as part of a conventional distance determination through a variogram-type interpretation. The assumptions around stationarity assume that for every pair of random variables $\{Z(x), Z(x+h)\}$ the covariance exists and depends on the distance of separation h. With the layered formation of the TMR over time, this distance h becomes too complicated to interpret.

The spatial relationship between samples is dependent on unnatural depositional processes caused by various human factors during the formation of the TMR. By way of illustration, an approach to determine the sample spacing for the following two TMR's, without taking into consideration the grade of the material would be different because of the creation and interpretation of what h would represent.

Figure 6.8: Genesis of two TMR's:
a) Left: Fine and coarse material combined and deposited from one central location
b) Right: Homogeneously sized material deposited in a time line

The Kimberley CTP TMR was created in a similar way to the TMR shown in Figure 6.8(b). This was taken into consideration during the determination of an optimal sample spacing.

In the case of TMR's, a natural model of horizontal codependence of grades does not exist and the formation of the TMR creates a pseudo randomised deposit where the layers follow the growth direction of the dump, having similar grades. The layers are created at the angle of repose of the material.

The Kimberley CTP fits the model described in Figure 6.9.

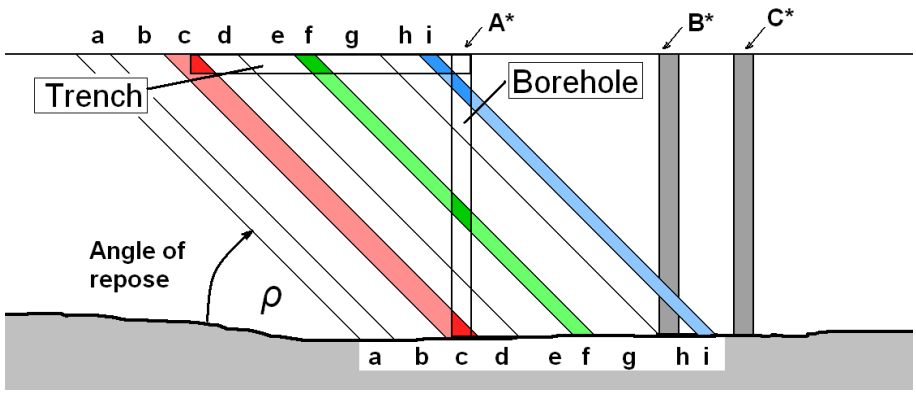


Figure 6.9: Layering of material and using the angle of repose (ρ) to determine sample spacing

Figure 6.9 is a schematic diagram illustrating the growth of the TMR in layers where a, b, c, etc.. are time slices over which material of similar characteristics was deposited. This implies that material that is deposited in period (a) has grades at the top of the TMR that will be the similar to the grade at the bottom. The same holds true for all the other periods ($b, c, d, e...$). Furthermore, if the size of the material going to the TMR is homogeneous (as is the case for the CTP), very little down slope segregation (in terms of particle size) takes place and the stone size and grade should remain homogenised within a specific period, from the top to bottom of the dump.

Ideally, a horizontal trench and a vertical drillhole will cross the same time-lines and should be representative of the same grade. In Figure 6.9 the top trench and drillhole represents the identical proportions of material if the angle of repose is 45^o.

As part of determining an optimum sample spacing, the horizontal distance between sample holes should allow for the intersection of each time slice at least once and there should be a nominal "time slice overlap" between adjacent holes.

In the case of the CTP TMR, the measured angle of repose averaged 34^o. The height of the TMR (in Figure 6.9) is 60m and applying Pythagoras' theorem, the sample spacing should not exceed:

$$tan(\text{Angle of repose of } 34^o) = \frac{60m \text{ height}}{\text{Sample spacing}}$$

$$\text{Sample spacing} \approx 89m$$

With this sample spacing, the different layers of the TMR will all be intersected along the direction of growth. Therefore $89m$ is the maximum distance that samples should be apart from one another.

After consideration of the requirement for local estimation and discussions between Dr. A Grills

(then De Beers: MRM department / currently Z Star Mineral Resource Consultants) and J Kilham (Kimberley Mines: Project geologist), it was agreed that a 60m spacing is appropriate. This will provide overlap from the top of one sample to the bottom of the next.

6.7 Local block estimation

Local block estimates are required for the purpose of mine planning and once the sample spacing was optimised, a method to create local estimates was investigated.

6.7.1 Estimating within homogeneous zones

The material that forms the TMR has periods of homogeneity in the same way as in-situ deposits does, excluding the effect of small scale random fluctuations. These periods are not necessarily only from homogeneous geological facies, but are better described over a time period as a chronofacies, with periods of similarity in geology and/or plant parameters. In the case of the CTP TMR, the initial chronofacies model proposed by Prins & Grills [56] is shown in Figure 6.10.

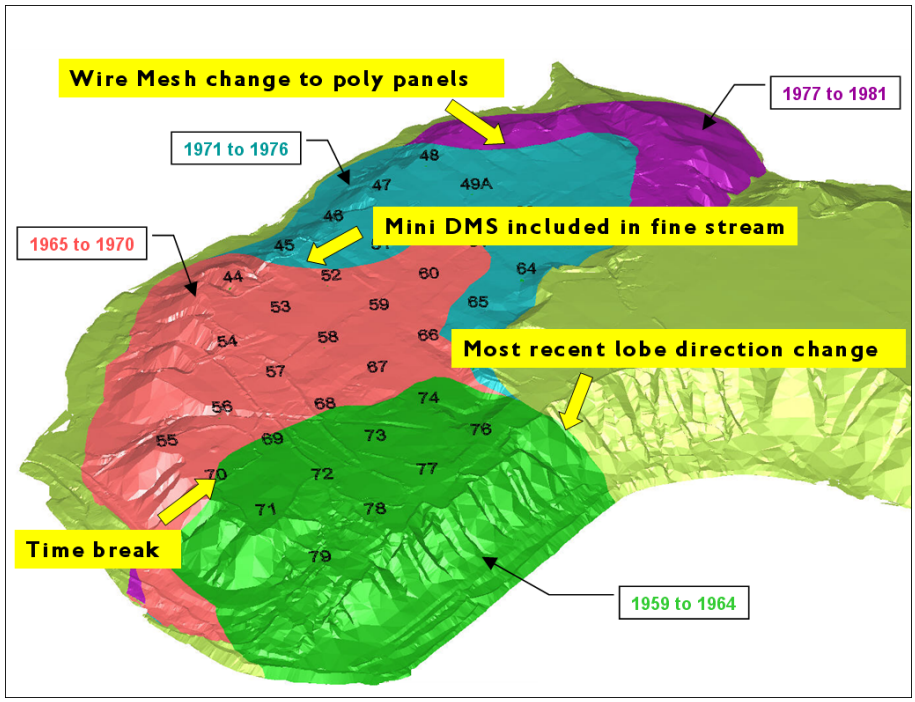


Figure 6.10: Initial chronofacies model (Interpretation: De Beers Kimberley Mines)

The colours represent different chronofacies with historical sample locations superimposed. An investigation into the grouping of the chronofacies was done.

The available data (albeit considered unreliable), when plotted as cumulative histograms (see Figure 6.11), showed that the first two facies (separated as a time break in Figure 6.10), have virtually identical cumulative distributions and can be grouped together for the purpose of

estimation, into a single chronofacies. Once grouped, the two chronofacies of 1959-1970 and 1971-1976 clearly have different cumulative distributions.

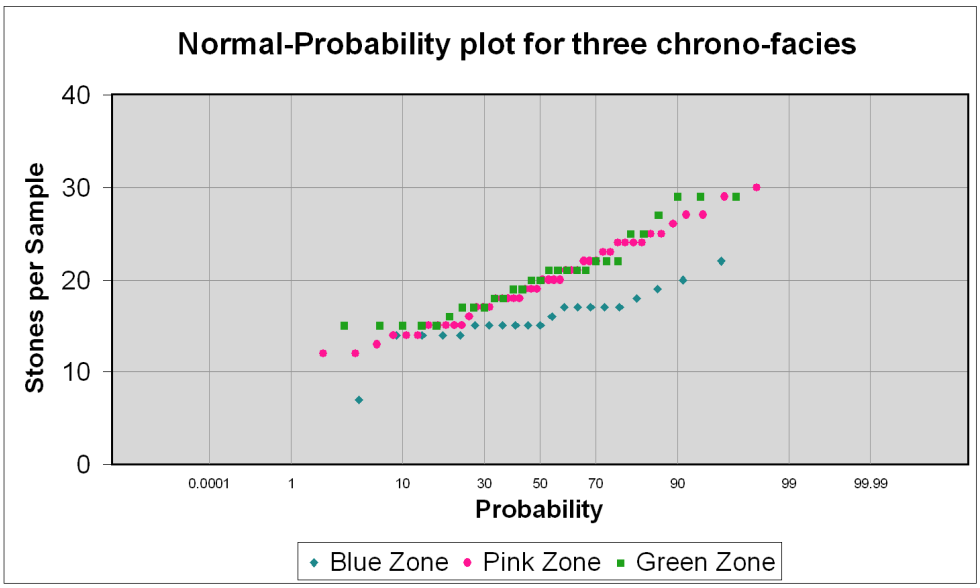


Figure 6.11: Comparing cumulative distribution functions between chronofacies

Records on the mine show that the following reduction in grade occurs along profiles drawn towards the chronofacies boundary (pink and blue zones):

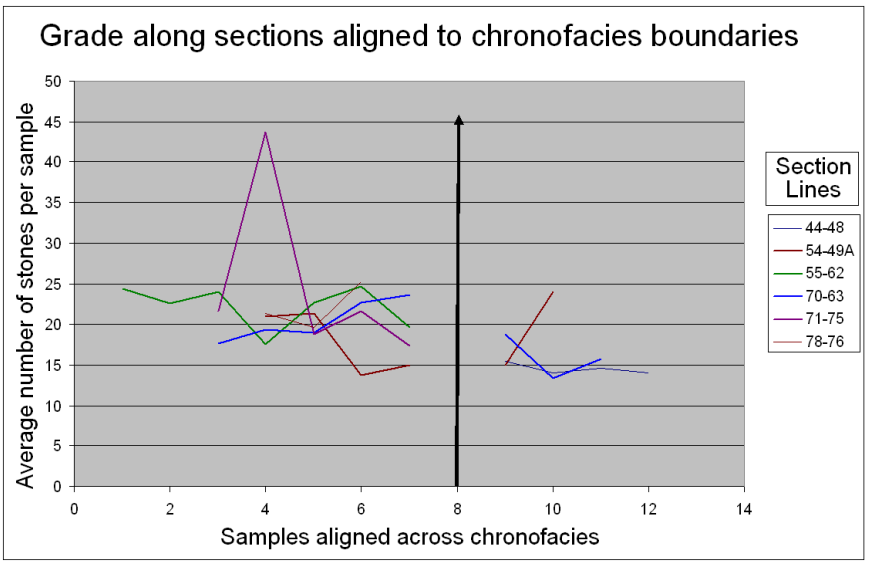


Figure 6.12: Change in average grade between two chronofacies (indicated by the vertical line): 21.5 to 16 stns/sample

A similar drop in grade was identified across a boundary related to the De Beers Mine, which was of slightly higher grade not being mined any more, and the grade on the TMR reducing accordingly. This change in average grade was used to introduce a new chronofacies.

The final chronofacies model as documented by Millad [50] is shown in Figure 6.13.

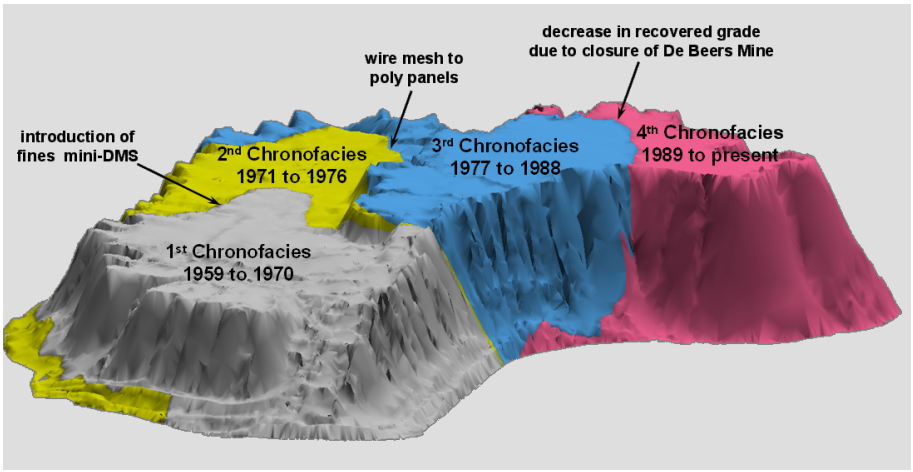


Figure 6.13: Final chronofacies used for the estimation of the TMR

Each of these facies was sampled and considered individually during the estimation process done my Millad [50].

6.7.2 Local block grade estimation

In a spatial evaluation process, samples inside the block are more representative of the variable being estimated and therefore should be given the most weight. When stationarity is assumed, more information is added from the neighbourhood through an acceptance of the codependence of values derived from the variogram. It is not possible to assume stationarity for TMR's and the proposed estimation method of kriging cannot be used. The estimation method applied is a weighted average of sample data.

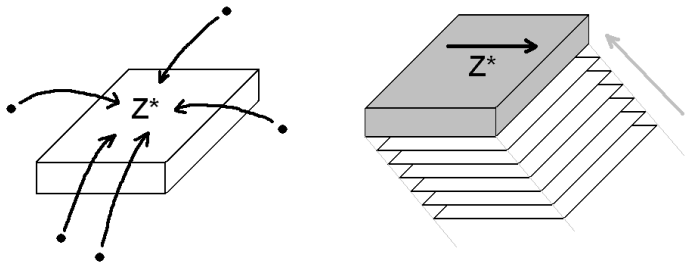


Figure 6.14: Kriging (left) vs. TMR estimation (right)

Figure 6.14 is a schematic showing the difference between kriging and the TMR estimation method.

Stationary deposits: (Figure 6.14(left)) The addition of information is via the framework of a geostatistical spatial model. For in-situ deposits, the variogram and change-of-support effect is used in a kriging process to calculate an estimate.

Tailings resources: (Figure 6.14(right)) The Kimberley CTP TMR comprises a succession of data collected along the growth direction and angle of repose. The samples are added and averaged to derive an estimate.

Approach: The estimated grade of a block is the weighted average of the grades based on the length of intersection of the samples with the block. The intersection between the sample and block is projected along the direction of the growth of the dump and the angle of repose. This is demonstrated in more detail in the next section and Figure 6.17.

With:

X = block to be estimated
L_i = length of sample s_i
l_i = length of intersection between projected block X and sample: $l_i = X \cap s_i$
s_i = sample$_i$
Z_i = grade in sample s_i

the estimator can be written as a weighted average grade:

$$Z^* = \sum_{\substack{\text{all} \\ \text{intersections}}} \left(\frac{l_i}{L_i} Z_i \right) \Big/ \sum_{\substack{\text{all} \\ \text{intersections}}} \left(\frac{l_i}{L_i} \right)$$

A block size of 120x120x10m was selected for the CTP TMR (discussions between Dr. A Grills & J Kilham), in keeping with the proposed mine plan and sample spacing of 60x60m.

6.7.2.1 Implementing the estimation methodology

For estimation, the different angles are defined to describe the formation and growth directions after examining the footprint and formation of the CTP and these include:
a) The angle of block grid network relative to the horizontal
b) The angle of (direction of) growth of the TMR and
c) The angle of repose.

An example of the angles are superimposed in Figure 6.15, from a presentation by Grills[32].

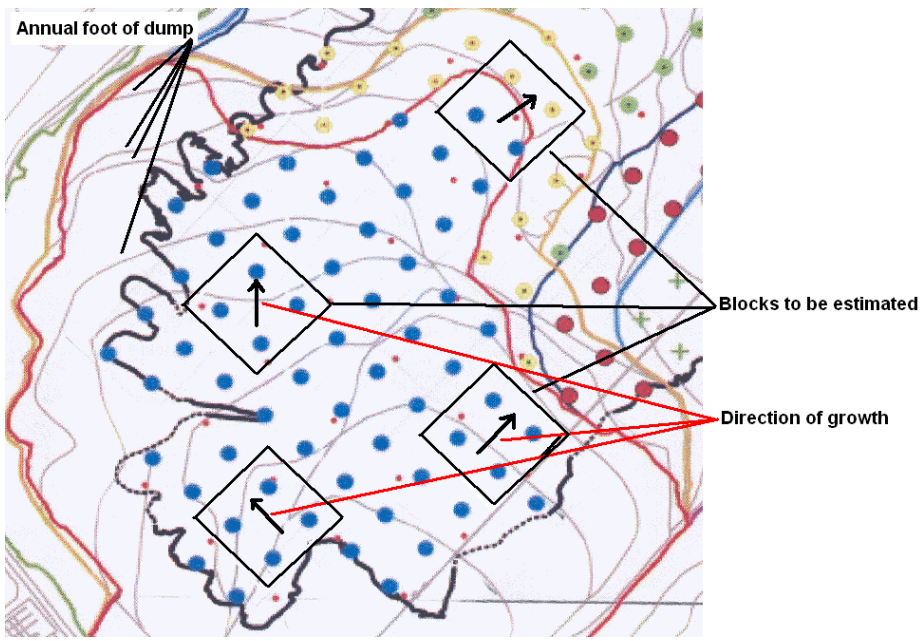


Figure 6.15: Example of growth direction of dump and block orientation

If the various angles are taken into consideration, the blocks and samples will interact with one another along the growth direction of the TMR at the repose angle. The illustration in Figure 6.16 shows the following block and growth direction interacting (from left to right):

a) Both samples intersect the block
b) Only the sample on the right will intersect the block and
c) Neither samples will intersect the block.

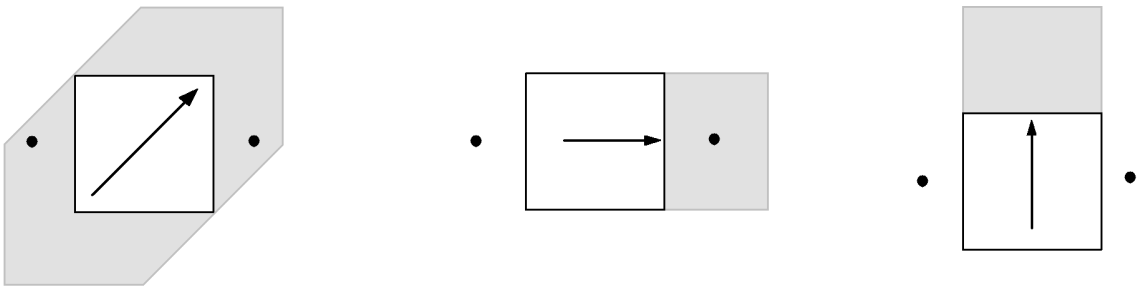


Figure 6.16: Growth direction (indicated by the arrows) and intersection between blocks and samples

The intersection of blocks by samples form the basis of the estimation method. It is equivalent to moving a block up and down along the repose angle from the top of the TMR to the bottom, and each section of sample intersected, is treated as a horizontal slice, added to a total and averaged to create an estimate for the block. Figure 6.17 illustrates this principle of intersection.

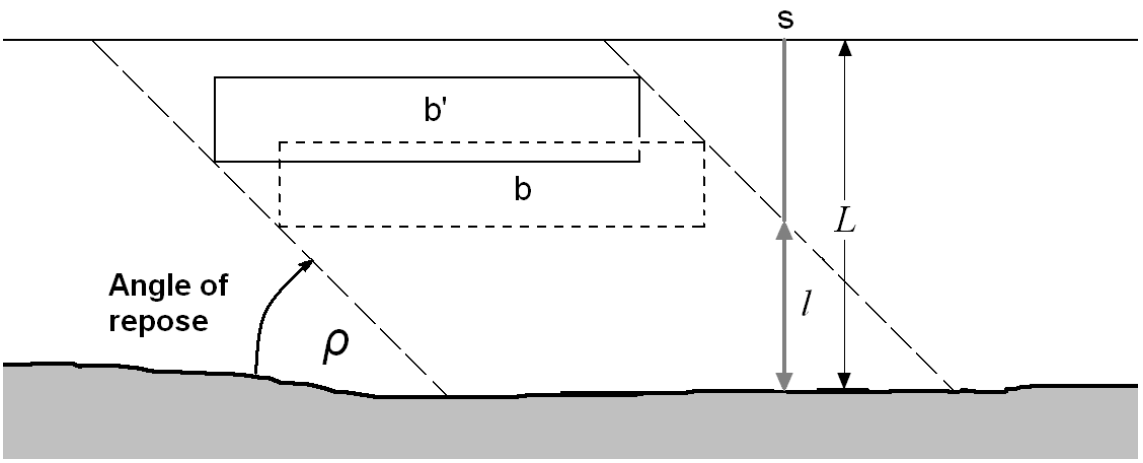


Figure 6.17: Angle of repose ρ and weighted intersection (l/L) between sample(s) and block(b')

Estimation methodology:

To facilitate software implementation, the samples were split into thin slices with thickness h and each slice assigned the measured grade associated with that length of sample.

If the angle of growth for the TMR (and shifting of the block) is defined as α, the translation of the block coordinates (x, y, z) by a horizontal distance r and downwards along the repose

angle by a vertical distance h is:

$$\begin{aligned} x' &= x + r\cos\alpha \\ y' &= y + r\sin\alpha \qquad \text{and} \\ z' &= z - h \end{aligned}$$

Figure 6.18 shows the translation of coordinates.

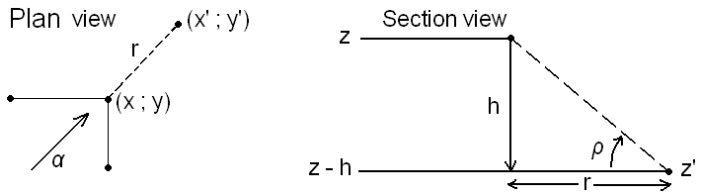


Figure 6.18: Translation of coordinates

The following relationship holds:

$$r = \frac{h}{\tan\rho}$$

the newly shifted coordinates for the translated block becomes:

$$\begin{aligned} x' &= x + \left(\frac{h}{\tan\rho}\right)\cos\alpha \\ y' &= y + \left(\frac{h}{\tan\rho}\right)\sin\alpha \qquad \text{and} \\ z' &= z - h \end{aligned}$$

All discretised samples falling within the shifted block are added to a running total to calculate the weighted mean that will become the block estimate Z^*.

6.8 Software Implementation

The software solution comprises two independent computer programs:

a) The first to enter and capture the direction in which the TMR formed and

b) The second program to calculate the estimates per block.

Figure 6.19 is an example of the first program used to capture the growth direction of the TMR.

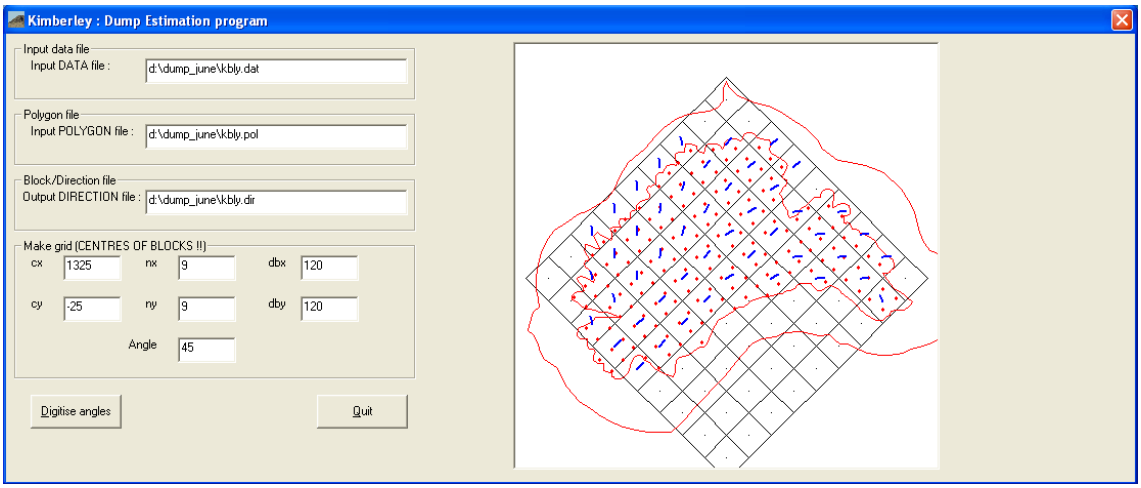


Figure 6.19: Digitising the direction of dump growth

Figure 6.20 is an example of the block evaluation showing the use of the growth direction and angle of repose of the TMR while creating the estimate.

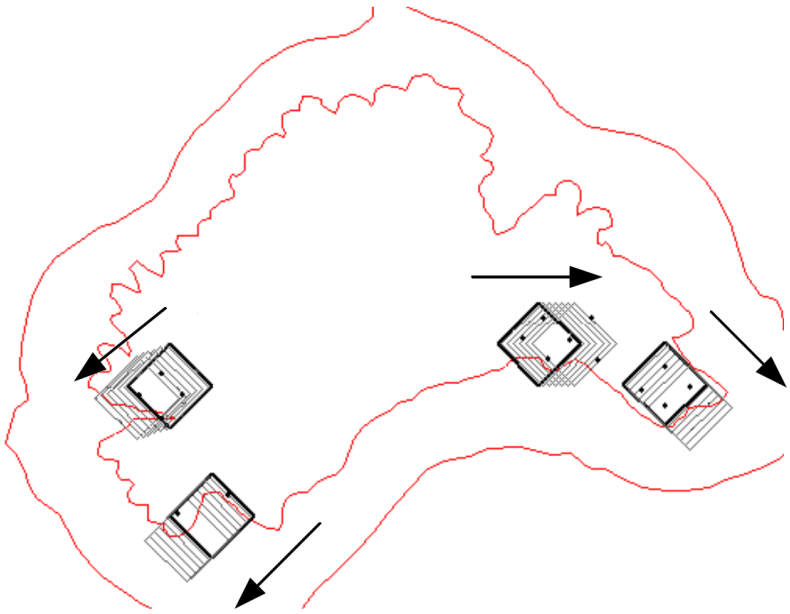


Figure 6.20: Adding up sample slices along repose angle and growth direction to create a block estimate

6.9 Determining an optimal sample size

The objective of determining a minimum sample size that is sufficiently large to express the measurement of grade within acceptable confidence limits is considered.

Generally for TMR's, historical data (sometimes lacking reliability) is the only source of information available and in some cases no measured grade information may be available, but a

sample size needs to be determined.

In the case of the CTP, an approach was needed to overcome the lack of reliability of the information. The economic break-even grade can be used as a minimum hurdle grade for the optimisation study. This hurdle grade is the lowest grade that will have a reasonable chance of making the mining financially successful. If the sampling process yields a grade less than this value, the TMR will be left unexploited.

Sample size studies for TMR's can be based on:

a) **Historical data:** A collection of sample datasets collected over an extended period, sometimes years apart, often with low levels of accuracy.

b) **An economic cut-off grade:** This grade value can be used as a minimum target value in the sampling optimisation. The finance department at mining operations will have some idea of what the economic break-even grade for TMR treatment should be. If the outcome of the sampling campaign is less than this grade, the TMR will not to be mined.

c) **Other TMR's:** When other TMR's are available with similar grade and characteristics, a comparison can be made with the outcome of their sample optimisation.

Although some historical data was available (i.e. Williams Auger), the use of an economic break-even grade was selected to determine sample size parameters for the CTP. The historical information was used for assurance and comparative purposes.

Objective: For the sampling of the TMR there are two sampling devices available, one with a diameter of 1.5m and the other with a diameter of 2.5m. The larger sample will produce a more representative sample, but will be more costly to operate. An analysis is required, based on minimal information, to investigate the effect of sample size on the confidence limits of estimates.

The basis on which the break-even analysis for the CTP TMR rests is a break even grade of 7.23cpht and an expected average stone size of 0.061cts/stn. Assuming a density of 1.9 tons/m^3, this is equivalent to having 2.25 stns/m^3.

Understanding the statistical behaviour of the TMR grade prior to the outcome of the sampling is speculative and in order to determine an appropriate sample size, different approaches and assumptions are applied in conjunction with consideration of the economic grade.

The idea is to assess and validate the choice of sample tool (1.5m versus 2.5m diameter) against a range of possible solutions and then to decide on the best "value for money" option.

Two different approaches were followed:

a) **A statistical model approach:** Different statistical models are applied to the grade measurements and assumptions are made about the expected behaviour of the sample results.

b) **Gy's approach:** This approach makes use of the fundamental sampling error of P. Gy's theory which is based on the physical properties of the material, as well as the expected grade.

Each approach has specific assumptions and requirements and produces different solutions. A decision is made once all the results are analysed to ensure that a minimum sample requirement is shared by all the approaches.

The two approaches are discussed in turn.

6.9.1 Using a statistical model approach

A range of statistical models are considered that represent reasonable assumptions about the grade and potential outcome of the sampling process. Applying these models will result in a number of solutions and the sample size considered sufficiently large will be chosen as the most conservative solution, i.e. the largest sample size.

In assessing the material for the CTP dump, grade and potential statistical models, the following assumptions are made:

a) All diamonds above the minimum cut-off size will be liberated in the mining process,
b) The material will generally be sized between $+1.5mm$ and $-9mm$
c) The Binomial, Poisson and Sichel models are the most likely statistical distribution models to occur.

These models have the following characteristics:

6.9.1.1 The Binomial distribution

If the material of the TMR is considered discrete particles with a probability of observing a diamond (p) and the probability for observing a non-diamond particles as $q = 1 - p$, the binomial distribution model can be applied to sample the population.

A schematic representation as documented by François-Bongarçon [25] illustrates the model with n particles taken as a subset (sample) of a much larger population N.

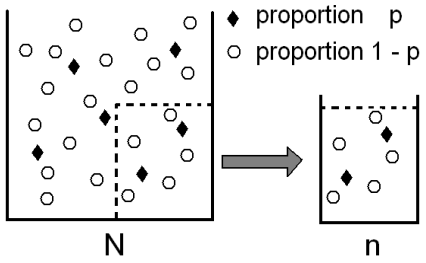


Figure 6.21: A binomial distribution model describing diamond occurrence

The statistical model as used by François-Bongarçon has a probability mass function:

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \qquad \begin{array}{l} x = 0, 1, 2, 3..., n \\ 0 < p < 1 \qquad and \\ q = 1 - p \end{array}$$

If X is a binomially distributed random variable, then the mean and variance of X are:

$$mean = \mu = np$$
$$variance = \sigma^2 = npq$$

The relative variance is defined as:

$$\text{Relative variance} = \frac{\sigma^2}{\mu^2} = \frac{npq}{(np)^2} = \frac{1-p}{np} = \frac{1}{\mu} - \frac{1}{n}$$

For diamond deposits where p is small and n is very large, the relative variance can be approximated by:

$$\text{Relative variance} \approx \frac{1}{\mu}$$

Furthermore, as in the case of diamond deposits, with $p \to 0$ and $n \to \infty$ and if $np \to \lambda$, the Binomial distribution approximates a Poisson distribution $P(\lambda)$.

6.9.1.2 The Poisson distribution

The assumption can be made that the sample will be taken large enough to contain many stones and the volume of the sample is the sum total of all the combined sub-samples. The stones per sample can then be modelled with the Poisson distribution to express the expected mean and variance of the estimate.

The method was described by Kleingeld [38] and uses the probability mass function for the Poisson distribution with $mean = variance = \lambda$ as:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, 3, ...$$

Consider the two sample tools (1.5m and 2.5m diameter) that can be used to take samples of 10m in length. The larger tool is more representative than the smaller tool, but is more expensive to operate. An assessment is required to test whether the smaller tool would recover sufficient numbers of stones to be representative and provide an estimate for a 120x120x10m block of the TMR within acceptable industry limits.

For the sample with diameter 1.5m: The dump is 60m in height and each 120x120x10m block will be estimated using approximately 24 samples. The volume of material (V) making up a local estimate would therefore be:

Volume used for estimating $Z^* = 24 \times \pi \times (1.5/2)^2 \times 10\ m^3 \approx 424m^3$

For this volume of material, at 2.25 stns/m^3, 955 stones are expected to be recovered.

How much confidence can be assigned to the recovery and use of this number of stones for an estimate of grade? Iterating the deviation from the mean in the expression of the Poisson confidence limits, with the aim to obtain 955 stones as an end result, result in considering a 6.34% deviation as follows:

Consider the volume of a sample as V with the probability of recovering a stone as p and modelling the number of stones per sample $n(V)$ as a Poisson distribution. With $n(V)$ large, this can be approximated by a Normal distribution with mean $= n(V)$ and variance $= n(V)$.

The confidence limits for a Poisson distribution with $E\{n(V)\} = pV = Var\{n(V)\}$, approximated by a Normal distribution becomes:

$$\begin{aligned} P\{-0.0634pV < n(V) - pV < 0.0634pV\} &= 0.95 \\ P\{\frac{-0.0634pV}{\sqrt{pV}} < \frac{n(V) - pV}{\sqrt{pV}} < \frac{0.0634pV}{\sqrt{pV}}\} &= 0.95 \quad \text{implies that} \\ G(0.0634\sqrt{pV}) - G(-0.0634\sqrt{pV}) &= 0.95 \\ 2G(0.0634\sqrt{pV}) - 1 &= 0.95 \quad \text{from which is follows that} \\ pV &= 955 \end{aligned}$$

Considering that on average 955 stones are expected to be used for the determination of the estimate Z^* of a block, the iterated confidence limit above shows that the estimate should be within 6.34% of the true value of the block in 95% of the cases. This is well within the acceptance of 15% deviation commonly quoted in industry.

This 6.34% variation in result is acceptable and thus the small sample tool with diameter of 1.5m is considered sufficiently large for use in sampling.

6.9.1.3 The Sichel distribution

Consider the same choice of two sample sizes as before: The smaller sample having a diameter of 1.5m and the larger a diameter of 2.5m, both with a length of 10m. The potential of obtaining zero stones per sample, by applying the Sichel distribution, as described by Sutherland [64], is used to investigate and express the sample size relative to the number of zero observations expected in the data.

Sichel's mixed Poisson probability distribution (with $\gamma = 1/2$ empirically determined by Sichel[60]) expresses the probability of recovering x diamonds in a sample with the probability density function:

$$P(X = x) = \sqrt{\frac{2\alpha}{\pi}} \exp\left(\alpha\sqrt{1 - 2\beta/\alpha}\right) \frac{\beta^x}{x!} K_{x-1/2}(\alpha)$$

with parameters $\alpha > 0$ and $\beta > 0$ and $K(.)$ the modified Bessel function of the second kind or order $1/2$.

The mean and variance of X is given by:

$$\begin{aligned} mean &= \frac{\beta}{\sqrt{1 - \frac{2\beta}{\alpha}}} \\ variance &= \beta \frac{1 - \frac{\beta}{\alpha}}{\left(1 - \frac{2\beta}{\alpha}\right)^{\frac{3}{2}}} \end{aligned}$$

and

$$\text{zero proportion} = e^{-\alpha\left(1-\sqrt{1-\frac{2\beta}{\alpha}}\right)}$$

The parameter β characterises the frequencies at the start of the distribution and when considering $2\beta/\alpha$, it characterises the tail of the distribution, with values close to unity implying a very long tail.

Depending on the sample size, but generally for small samples in a low grade environment, the typical shape of the distribution is J-shaped with a long upper tail.

The parameters for the distribution can be obtained by using the zero proportion $\phi(0)$ of the data and the mean $\overline{x}$. The parameters α and β can be determined from:

$$\alpha = [-\ln\ \phi(0)]\left[1 - \frac{ln\ \phi(0)}{2\ [\overline{x}\ +\ ln\ \phi(0)]}\right] \qquad equation(1)$$

and $$\beta = \frac{\overline{x}}{\alpha}(\sqrt{\overline{x}^2 + \alpha^2} - \overline{x}) \qquad equation(2)$$

To calculate the probability mass function, a recurrence formula can be used as follows:

$$\phi(0) = exp[-\alpha(1 - \sqrt{1 - 2\beta/\alpha})] \qquad equation(3)$$
$$\phi(1) = \beta\ \phi(0)$$
$$\phi(x) = \frac{\beta}{\alpha}\left[\frac{2x\ -\ 3}{x}\right]\phi(x-1) + \frac{\beta^2}{x(x-1)}\ \phi(x-2) \qquad x = 2, 3, 4, \ldots$$

This distribution is typically used to model data obtained from areas in which diamonds occur as clusters in a fairly low grade deposit, and which has been sampled with a relatively small sample tool.

Application to the CTP TMR

For the determination of sample size prior to the beginning of the sampling, Sutherland uses the Sichel distribution to make the zero proportion the stated aim of the sampling programme.

As explained by Sutherland[64]:

Assume for example a grade of 1 stn/m^3, the average number of stones per unit sample for $\overline{x}$ is one. If the sample size is 5m^3, the value of $\overline{x}$ is 5 and so on. With a declared aim of achieving a specific $\phi(0)$ value and with knowledge about $\overline{x}$, the sample size can be varied iteratively in equations 1, 2 and 3 above to establish the minimum sample size that corresponds to the desired $\phi(0)$ value.

The minimum sample size determined in this way is used as the most cost efficient sample size that fulfils the statistical requirement of being a Sichel distributed variable, with a stated probability not to exceed a threshold proportion of barren samples.

Application to the CTP:

This method is applied to the CTP as follows:

For a stated aim that no more than 0.5% of the samples will be accepted as zero values, how big a sample is required?

As per Sutherland the average grade ($\overline{x} = 2.25 stns/m^3$) is split into two components:
a) volume per sample (m^3) and
b) stones per sample.

The volume of the sample is minimised and α and β determined from the Sichel distribution such that:
a) stones/m^3 = 2.25
b) zero proportion = 0.005
c) $\alpha > 0$ and $\beta > 0$
d) $exp[-\alpha(1 - \sqrt{1 - 2\beta/\alpha})]$ remains equal to 0.005

Multiple solutions exists for α and β. Take for example α: solutions exists as long as $\overline{x} + ln\ \phi(0) > 0$ (from equation 1). A unique minimum sample size however exists and was determined as 2.35m^3.

If each block is sampled by four 10m samples with diameter 1.5m, and according to the estimation method, the block is on average intersecting twenty four samples along the growth direction and angle of repose when determining the block estimate Z^*, the total volume of material that would be used for the local estimate would be 424m^3. This volume of material is many orders of magnitude larger than the minimum determined sample size using Sichel's distribution. Even the individual sub-samples with a volume of 17.67m^3 is many times larger than 2.35m^3.

As a result, the probability of observing zero stones in a 424m^3 sample by assuming the Sichel distribution is much less than 0.5%.

6.9.2 P. Gy's fundamental sampling error

Pierre Gy's sampling formula (appendix E and P. Gy [33]) as interpreted by Royle [59] is expressed as:

$$\sigma_e^2 \ = \ \left(\frac{1}{M_{sample}} - \frac{1}{M_{Lot}}\right) \ c \ f \ l \ g \ d^3 \ A^2$$

The constants for f (shape), g (granulometry), and l (liberation) can be determined from visual observations and metallurgical knowledge of the TMR material. Since diamonds occur as a rare mineral, the simplified expression of c is used.

$$c \approx \frac{D_m}{a_L}$$

with
D_m: Volumetric density of diamonds (3.52 $g.cm^{-3}$)
a_L: Decimal proportion of mineral A (the diamonds) in the lot L (as a proportion).

The value for d (particle size) is taken from measurements of the diamond size frequency population and A (the grade) is based on the financially determined break-even grade.

Research by François-Bongarçon [26] enhances the application of Gy's formula by refining the liberation factor as $l = \left({}^{d_l}/_{d}\right)^b$, with d the size of fragments, d_l the maximum size of pure "metal" grains (diamonds) and b is an additional parameter that can be adjusted based on experimental results. For a once off application on the CTP TMR, the parameter b could not be calibrated and the liberation factor was set as $l = 1$.

Parameterisation for the CTP TMR

For the CTP TMR, the break even grade is 7.23 cpht. Assume a density for the non-diamondiferous (gangue) material of $D_G = 1.9$. The density of diamond is $D_m = 3.52$ and with 1 ct = 0.2g, the grade can be expressed as:

$$\begin{aligned} \text{Grade= A} &= 7.23\ cpht \\ &= 1.446g/52.63\ m^3 \\ &= 0.027\ 474\ g/m^3 \end{aligned}$$

Furthermore,

$$\begin{aligned} \text{Proportion} = a_L &= 7.23\ cpht \\ &= 7.23 * 0.2/(10^8)\ \ gram/gram \\ &= 1.446 * 10^{-8}\ \ gram/gram \end{aligned}$$

Using the size frequency information as documented by Millad [50] of the historical TMR samples shows that the 95% percentile for the log-normal size distribution is $p_{95} = 2$ cts. This equates to an opening of around 5.56mm, thus $d = 0.556cm$.

For the configuration of the drilling equipment, two possible sample sizes for M_{sample} were considered:

a) A drill bit with diameter 1.5m (weight = 33.6 tons for a 10m lift) and
b) A larger drill bit with diameter 2.5m (weight = 93.3 tons for a 10m lift)

As for the distributional approach, taking the angle of repose into consideration, a total of around 24 samples represent the total weight of material used to estimate the grade (Z^*) of each block.

Applying Gy's formula for samples with diameter 1.5m and height of 10m

$$\begin{aligned} \sigma_e^2 &= \left(\frac{1}{33.575 * 24 * 10^6} - \frac{1}{273\ 600 * 10^6}\right)\ (3.52)\ (0.5)\ (0.25)\ (1)\ (0.556)^3\ (0.027\ 474)^2 \\ &= 0.000\ 004\ 88\ (g/m^3)^2 \end{aligned}$$

For confidence limits around the targeted grade of 7.23cpht and transforming units of g/m^3 back to cpht gives:

$$1.96\sqrt{\sigma^2} = 1.16 \text{ cpht with confidence limits}$$
$$\text{Grade conf. limits}_{95\%}: \quad [\ 6.07\ ;\ 8.39\ ]$$

Applying Gy's formula for samples with diameter 2.5m and height of 10m

$$\sigma_e^2 = \left(\frac{1}{93.266 * 24 * 10^6} - \frac{1}{273\ 600 * 10^6}\right)(3.52)\ (0.5)\ (0.25)\ (1)\ (0.556)^3\ (0.027\ 474)^2$$
$$= 0.000\ 001\ 75\ (g/m^3)^2$$

For confidence limits around the targeted grade of 7.23cpht and transforming units of g/m^3 back to cpht gives:

$$1.96\sqrt{\sigma^2} = 0.696 \text{ cpht with confidence limits}$$
$$\text{Grade conf. limits}_{95\%}: \quad [\ 6.53\ ;\ 7.93\ ]$$

Summary of confidence limits

Table 6.3: 95% Confidence limits around the break-even grade

Sample Size (x10m length)	Lower Conf. limit	Economical cut-off grade	Upper Conf. limit	% variation to mean
Diameter 1.5m	6.067	7.23	8.393	16.1%
Diameter 2.5m	6.534	7.23	7.926	9.6%

In summary, samples with a diameter of 1.5m and length of 10m are acceptable based on the fact that they are expected to produce an average grade estimate Z^* within 16% of the true grade in 95% of the cases. If the true grade of the CTP TMR is higher than the break-even grade, the level of confidence can be stated with even more confidence. This level of variation is considered acceptable.

6.10 Practical implementation

During this study it became evident that many "unknown" factors, experimental arguments and specific situations can be postulated for which the methodology will have varying degrees of success. The correct statistical model can only be confirmed once the data has been collected, but by then the sampling has been completed and it is too late to make any modifications to the sampling parameters. Since the CTP TMR's layering and small scale behaviour has not been mapped and remains unknown (as with almost all other TMR's in De Beers), it can broadly be stated that TMR sampling and estimation can never be considered as an exact science.

Each TMR is unique in it's creation and character and should be assessed independently, but the approach proposed for the CTP TMR forms the basis for designing appropriate sampling and evaluation programmes for other TMR's.

A short summary is given demonstrating arguments and pictures of what should be considerations when sampling TMR's.

Segregation

TMR material is homogenised in terms of particle size through crushing and sieving and segregation should not occur. Mass flow processes govern the deposition of the tailings on the dump, but when this assumption does not hold, segregation should be taken into account.

A schematic is shown in Figure 6.22, showing segregation for material when consisting of a wide range of particle sizes. The larger and heavier particles segregate towards the outer limits of the deposition.

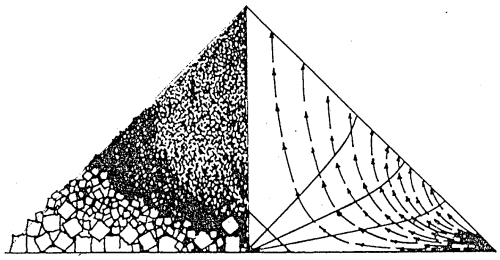


Figure 6.22: Segregation model for heterogeneously sized material

When the particles are of a similar size, the material flows downwards along the side of the TMR. Figure 6.23 shows such a mass flow at the Koffiefontein TMR on which no segregation was observed.

Figure 6.23: Material "flowing" downwards along the side of the TMR

Assessing the homogenisation of the material going through the recovery processes that supply the CTP TMR, no segregation of material is expected.

Overlapping TMR material from different lobes

Detailed inspection of results and application of the estimation methodology must be done on a block by block basis when the dumping of material overlaps from different lobes and directions, as shown in Figure 6.24.

Figure 6.24: Overlapping dump growth (Example: Namdeb #4 plant)

The extrapolation of borehole information along the repose angle and growth direction of the dump must consider the material relevant to the block being estimated.

In Figure 6.25 (left), the borehole on the left must be used with a growth direction from the left and (right) the borehole on the right must be used with growth direction from the right.

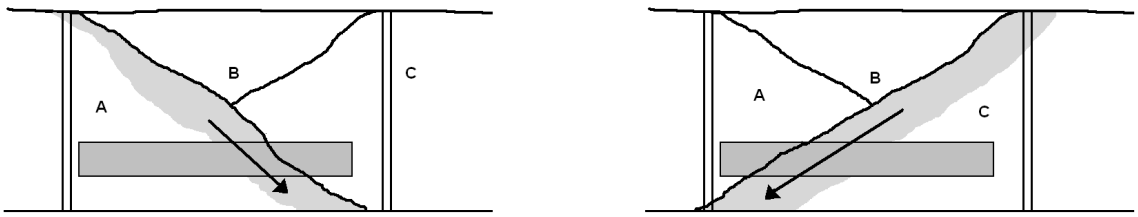


Figure 6.25: Different overlapping directions relative to the appropriate extrapolation of sample information

The overlapping of lobes happen frequently and must be dealt with accordingly by splitting up the chronofacies model, data and growth directions used in the estimation process.

Dealing with surface enrichment

The material covering the outer layer of the dump must not be included in the sampling because it is not representative of the underlying material. This material could potentially be subjected to a process of surface enrichment, where weathering could have broken down the kimberlite or finer material may have been removed by weathering processes. This will leave behind coarse and potentially heavier/larger diamond particles that can bias the sample results upwards.

Sample material when stored on the surface of the dump must take this into consideration.

Figure 6.26 shows an example of coarse material on a TMR in Kimberley, where potential surface enrichment could occur.

Figure 6.26: Surface enrichment (Example: Kimberley Stadium TMR)

Precautions were taken during the sampling of the Kimberley CTP TMR to avoid the sampling of potentially enriched surface material. Also, where the samples were stored, the surface material was removed to avoid contamination of the sampled material.

Figure 6.27 shows the precautions taken by bulldozing the surface material away from the sampling and storage site.

Figure 6.27: Sample storage site preparation (Example: Kimberley CTP dump)

The Google Earth picture in Figure 6.28 shows the status of the sampling (May 2004) and storage of samples on the TMR.

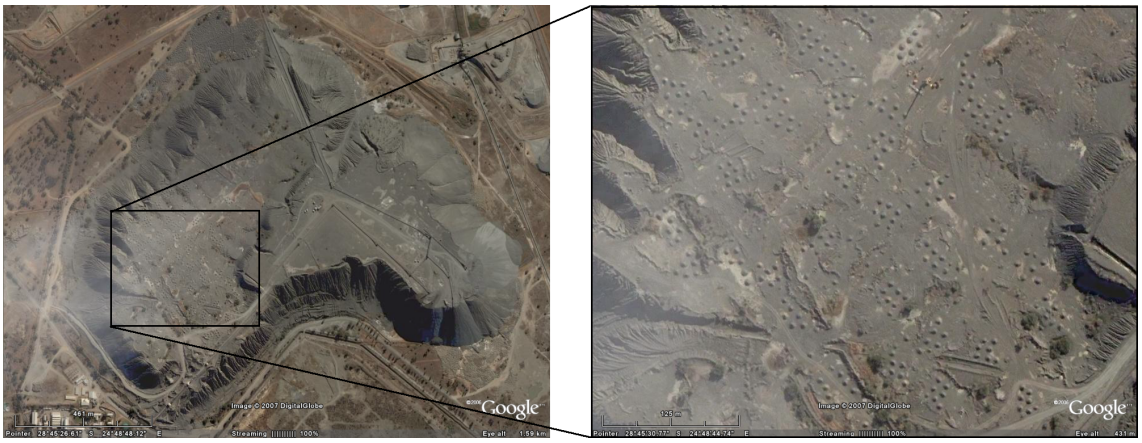


Figure 6.28: Kimberley CTP TMR (left) and a zoomed in area (right) showing more details of the sampling activities

Figure 6.29 shows samples stored on the TMR awaiting treatment and also showing the removal of material to avoid potential surface enrichment.

Figure 6.29: Six samples per drill hole (of 10m length each), stored on the dump and awaiting treatment

Using an initial wide grid for sampling

If it is operationally possible and if the drilling equipment is sufficiently mobile and not restricted by infrastructure on the dump, a wide spaced sample campaign (skipping every alternate sample of the proposed grid) could be drilled as a first phase. Rapid treatment of the sampled material can give an indication whether the global grade will be high enough to continue detailed sampling of the TMR, or whether further sampling should be terminated.

This strategy will also enable the testing of the statistical models and remedial steps can be taken timeously if modifications are required to improve the sample optimisation.

For the CTP this was not required, as historical information existed that could be used for validation of results and the outcome of the sample results was according to expectation.

6.11 Conclusion

The challenge of sampling the CTP dump is that the information available from previous sampling campaigns was referred to in reports as unreliable. The sample optimisation could therefore not be based on this data and thus the concept of using the minimum economic cut-off grade was introduced as a target grade to achieve in the sample optimisation. Similarly, the distribution of the sample grade is not known, but a model based approach using the Poisson and Sichel distributions, as well as the sampling theory as developed by P. Gy was used to determine a range of solutions.

The Sichel distribution showed that with a long tail for the distribution and using a very low acceptable zero proportion that a sample size with 1.5m diameter and 10m length will always contain some diamonds. The total volume of material used for the estimation of the block (33.6 tons) will therefore be sufficiently informed for a reliable estimate.

With enough diamonds present in the sampled material used for the estimate, the Poisson distribution models the variation in the estimate as 6.34%. This is much less than the industry benchmark of 15%. A sample grid of 60x60m grid is proposed with sample heights of 10m.

For determining an optimal sample size, statistical models using the Poisson and Sichel's distributions, as well as P. Gy's sampling formula, were successfully implemented to establish that a sample size using a 1.5m diameter x 10m high sample would recover sufficient numbers of stones to be considered representative.

The method developed for doing local block estimates was successfully implemented by Millad [50] to the chronofacies I of the Kimberley CTP tailing mineral resource.

The techniques and approach to estimates developed in this research, had also been applied successfully to other TMR's of De Beers.

(Deliberately left blank)

Conclusion and perspectives

The topics covered in this thesis are varied and cover challenges in secondary diamond deposits of sampling and estimation, as well the parameterisation of primary kimberlites exploration targets and the analysis of indicator minerals collected during the exploration process.

Diamond deposits (and especially secondary diamond deposits) are recognised as being difficult to estimate relative to other commodities and as new areas are identified for exploration and subsequent evaluation, new methods of analysis, or innovative application of existing techniques are required to deal with various challenges encountered.

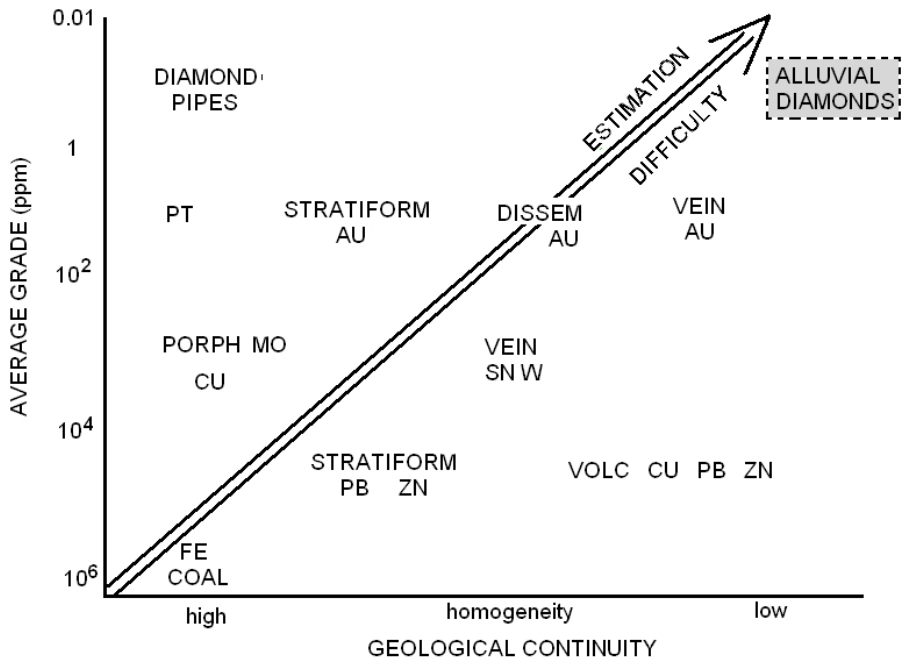


Figure 6.30: Relative difficulty to estimate different commodities (after King et al., 1982)

Techniques applied in this thesis such as data mining (classification trees), probability models and geostatistics are well founded and provide results used for planning and decision making on a daily basis. Early on in resource development, without enough grade data available, models are speculative and it is in such situations (as in chapter 4) where it becomes difficult to make decisions and costly if mistakes are made. Any guidance that can be given and a "more-correct-way-to-proceed" is the best that can be done until some data becomes available.

However, when data is collected, the biggest challenge lies in the quality of the data. The challenges described in the introduction chapter are always present to a more or lesser extent when evaluations are done and needs to be dealt with.

In many cases, the combined effect of all the estimation challenges encountered in the development of a resource are factored out of the equation when resource statements becomes reserve statements, when expressing what could reasonably be expected for extraction.

These factors are built up of many components which one should not become complacent about, as improvements along the resource development process can have significant financial implications in the end.

It remains important to have an accurate estimate for the resource, against which most efficiencies are measured. Statistical techniques used to do the resource estimates have been well researched, which leaves the single biggest unresolved issue that of representative sampling.

Non-representative sampling, due to the discrete nature of diamonds can easily happen. This is a field of research where improvements can lead to more accurate estimates and reduced cost - exactly what all company objectives aim to achieve.

This thesis addressed some issues toward achieving this, but due to the ever changing nature of the challenges facing secondary diamond deposit evaluation, the need for research is ongoing and there will always be opportunities to improve. The most relevant perspectives on how to further this research can be summarised:

a) The Nearest Neighbour and Gabriel Graph methods enabled the progressive creation of clusters as link distances changed, with associated average cluster size. An improvement in this regard would be to investigate the effect of introducing directional components into the analysis.

b) When analysing indicator minerals (IM's), the introduction and testing for directional components based on the regional movement of the particles by erosional forces would improve the classification tree models and the choice of distances for the classes, different from the current model's (5, 10 and 20km) classes should also be investigated further.

c) The use of the Cox process for creating diamond simulations is well established in sample optimisation studies. Using such simulations, research showed that 7m^2 is the minimum sample support to be used, with 10m^2 preferable for marine deposits. This research was parameterised for higher grade areas than those planned for future mining, which intuitively suggests that larger samples will be required in time. Since the taking of even larger samples than the current 10m^2 samples is not possible (from an engineering and cost perspective), future research will focus on the clustering of samples. Specific attention will have to be given to research the creation of the simulation to express a realistic behaviour, similar to the diamond resource, at those distances over which the clustering will be done.

d) In future, technology currently being tested will improve the mapping of entrapment features in shallow water marine deposits. Since the presence of a trap is not an absolute guarantee that diamonds are present, research will be required to ensure that knowledge of the entrapment information is used responsibly to establish where diamonds are present, but simultaneously, when a trap is sampled, to find methods to adjust the estimates made for the mining panels to take cognisance of the bias created by the sample strategy. The ingenious use of geological

interpretation, sample spacing and size, as well as an appropriate estimation method would be required to achieve this.

Taking cognisance of these perspectives and with the continuation of research in the pursuit of excellence will ensure the successful mining of secondary diamond deposits.

Although the thesis describes diamond specific problems and case studies, slight adaptations will make methods and results applicable to other minerals as well.

* * * † * * *

(Deliberately left blank)

Appendices

(Deliberately left blank)

A The impact of changing sample size and spacing, expressed as a return on investment

The following was prepared for an internal report[3]:

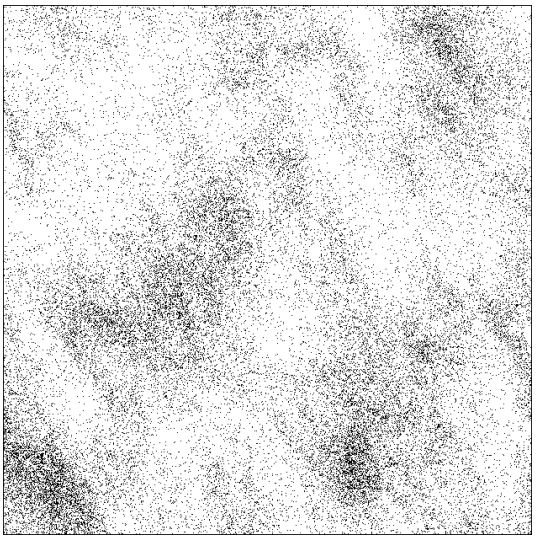
Expression of Value-add by Means of NPV, for the 2-D Non Stationary Deposit Research Project

Abstract: The 2-D non-stationary project's primary focus lies with the development of mathematical and/or statistical methods to make the analysis for 2-D deposits scientifically and financially robust. A request was made to express the financial value-add of this research and an exercise was done to show how, through using statistical developments, financial benefit can be derived.

Introduction

Questions are often asked around the optimal use of samples and sampling equipment. A typical question would be: "How many samples must we take, how big should they be and how far apart must they be?"

This is an easy question to ask and many statistical tools exist to help answer these types of questions, but one must reflect upon the complexity that quite often, either no information exists, or very little and often incorrect or incomplete information exists on which the proposed calculations are based. In answering such questions objectively, scientific methods are used, often based on the outcome of a geostatistical simulation exercise. In short, such an exercise involves the creation of a pseudo-realistic deposit of stones deposited on the seabed - see the example right. This deposit has statistical and geostatistical properties similar to that of an area targeted for sampling and estimation.

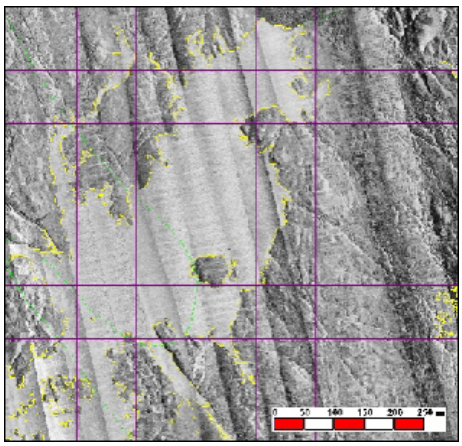


For all practical purposes, this "deposit" can now be sampled and evaluated just as would be the case in reality - without any expenses incurred. Such an exercise can indicate what sampling and evaluation strategy is best suited and what potential returns may be realized.

[3] *Expression of Value-add by Means of NPV, for the 2-D Nonstationary Deposit Research Project* by Chris Prins, De Beers Internal Report, 30 September 2005

Technology advance through research

Some of the deposits on the west-coast are spatially more complex than shown in the figure above and the deposits are becoming more marginal and less money is being made available for sampling. More effort must therefore be put into the understanding of the sampling and evaluation to ensure correct mining or walk-away strategies. An example of such a deposit is shown (right).

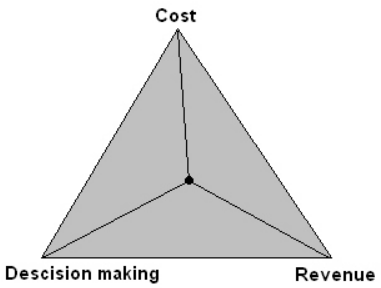


In these deposits, the direction of mineralisation is not constant and varies according to local geological features. The research is aimed at capturing this variation in the models to improve the decision making for sampling and estimation in more complex west-coast type deposits.

Example

The gains that can result from such an analysis and research is illustrated as a combination of three factors: cost, revenue and decision [4].

The full benefit derived from the proposed research is a combination of the three factors. In certain circumstances, cost may not be reduced, but the decision making will be more robust. In such a case it is more difficult to express the financial benefit, but the example below, based on actual data, will show how better decision making influences the financial evaluation of a deposit.



The example is based on a part of De Beers Marine Namibia's Atlantic I resource off the west-coast of South Africa. The research done in 1996 was to determine what size sample tool would be sufficiently large for successful and representative sampling.

At that stage, the appropriate technology was developed and an optimal sample size was determined. Over the past number of years, the financial gain was realized, but never expressed nor linked to the contribution the scientific research process made. This example will attempt to do so.

A scenario with two options were analysed:

a) Sample an area with the Mega-drill (0.72m^2) on a 33.3x30m grid in a cluster of 3 and
b) Sample the same area with the Deca-drill (10m^2) on a 50x50m grid.

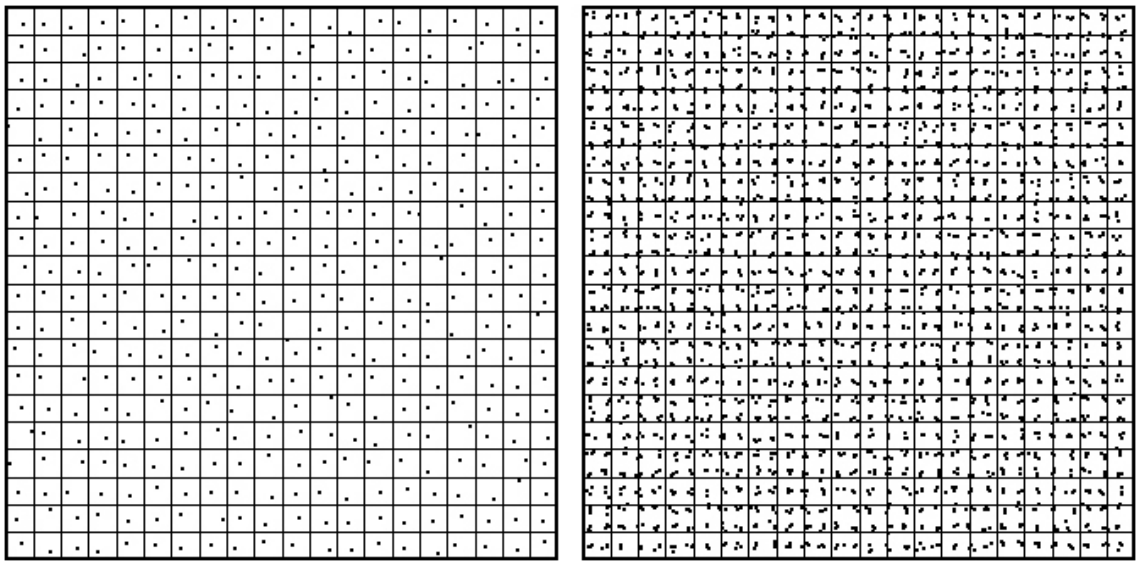
What would the financial difference be between the two approaches to sampling and evaluation?

The simulation shown in the introductory paragraph was sampled 100 times, each in turn with

[4]Source : Internal De Beers presentation: Coward & Nicholas, 2005

the sample sizes 0.72m^2 and 10m^2 as is the practice in De Beers Marine, i.e. 50x50m spacing (10m^2) and 33.3x30m spacing (clusters of three samples, each 3*0.72m^2 = 2.16m^2).

Illustratively:

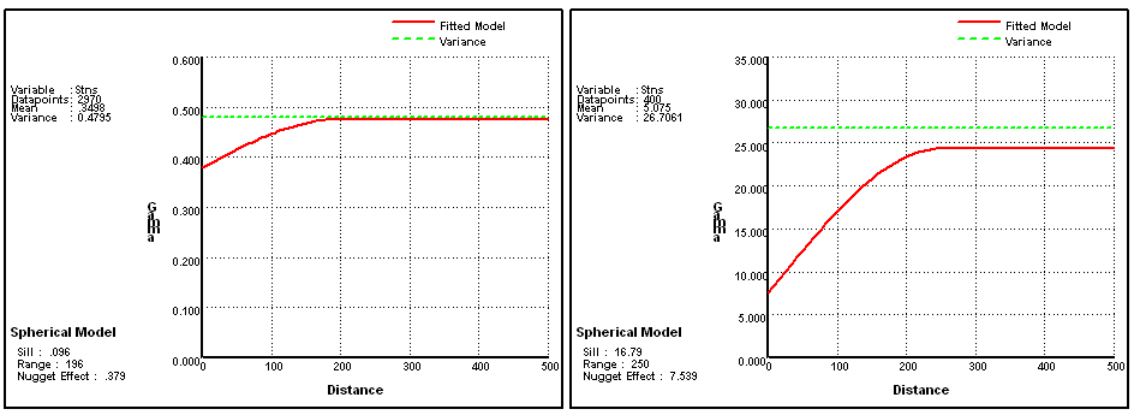


From these 100 sampling campaigns, a subset was selected for each of the sample sizes at random and the variograms calculated. An average variogram was determined for each sample size by averaging the variogram parameters.

Results were:

$$10\text{m}^2 : \gamma(h) = 7.54 + \text{Spherical(Sill=16.79, Range=250)}$$

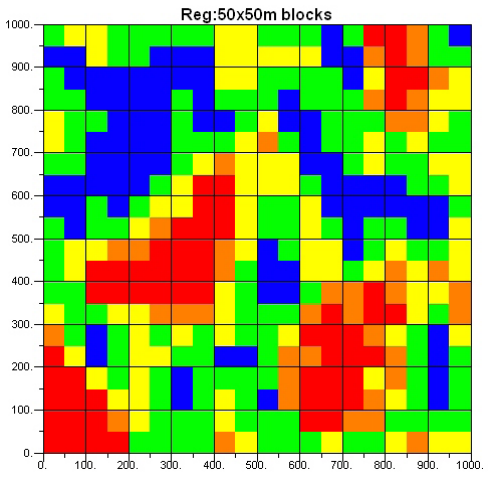
$$0.72\text{m}^2 : \gamma(h) = 0.380 + \text{Spherical(Sill=0.096, Range=196)}$$



These variograms were used to krige the 100 outcomes, for each of the two sampling campaigns.

The "true" number of stones per 50x50m block was calculated and graphically the results are as follows:

For each of the 50x50m blocks in the graph(right), there are 200 estimated values. One hundred generated on 0.72m^2 samples using kriging methodology and one hundred values based on 10m^2 samples, also using kriging.

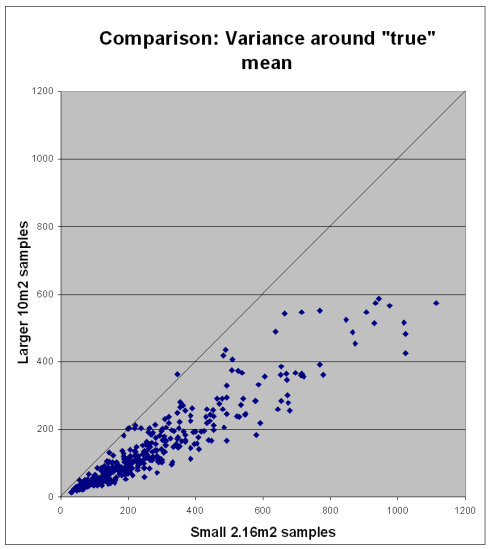


A comparison was made between the two sets of results to see which set has the lowest variance around the true grade. Each of the block comparisons were standardised relative to the "actual" value of the block and expressed as a percentage by:

$$\%\text{diff} = \left(\frac{\text{actual} - \text{estimated}}{\text{actual}}\right) * 100$$

Through this method of comparison, the better decision would be to choose that sample campaign that gives the lowest variance - in other words the "highest probability" of being correct.

Looking at the variance of the relative differences, a comparison can be made between the variability of the results of the 10m^2 samples and the 2.16m^2 samples. The campaign with the least variance is the better campaign and expresses a higher financial confidence in the estimate. Displaying these results as a X-Y plot is as follows (see right). It is clear that the 2.16m^2 samples are much more variable than the 10m^2 samples, through the comparison to the 45^o slope of the line and $\max_{2.16m^2} = 1100$ vs. $\max_{10m^2} = 600$.



Comparing the average variance of the two sets:

2.16m^2 : $\sigma^2 = 270$

10m^2 : $\sigma^2 = 143$

If the ratio 270/143 is used, it reflects that the 2.16m^2 samples have 88% more variation in results than the 10m^2 samples.

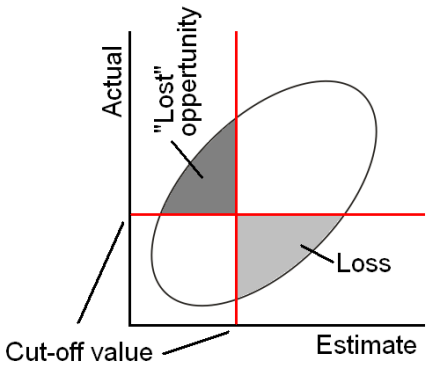
The choice of sample campaign would be easy to make (for simplicity, ignoring the cost of sampling for the time being): The sampling campaign using the 10m^2 samples is more acceptable and there should be a financial reward for making this choice, both financially as well as by reduction of the discount factor (risk factor) used in the net present value (NPV) calculations.

Financial impact

Once the decision has been made to sample with the 10m^2 samples, how much difference has this decision made financially?

This question is answered relative to a financial break-even value. For the purpose of this analysis, different cut-off values are shown, to express the variability that can be expected.

Consider the following figure:



If the two red lines represent the economical break even value, the grey shaded areas are either not mined ("lost" opportunity) or mined, but at a loss ("loss").

As the cut-off changes from low to high grade values, the relative values and percentages of "loss" and "lost opportunity" will also change. For example, at a cut-off of 300 stns/block the following will result.

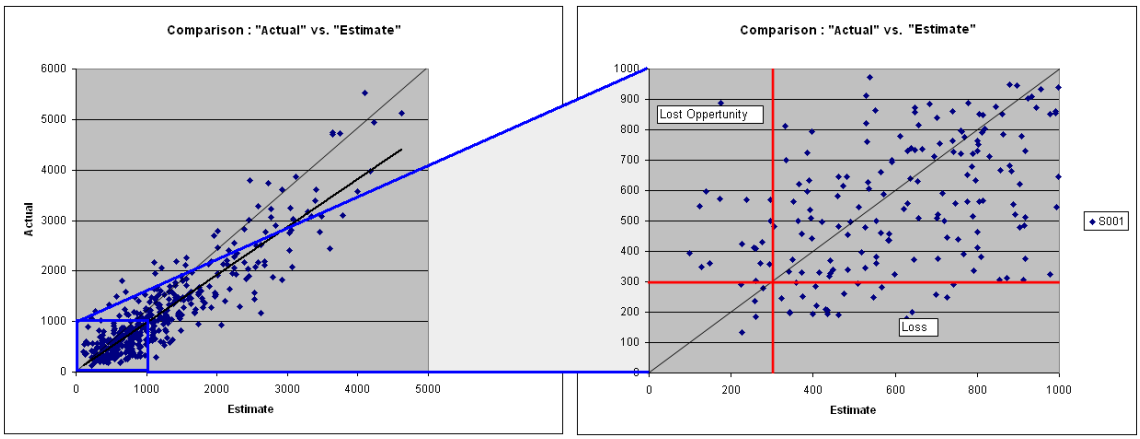


Figure 6.31: (Left) Comparing "Actual" against "Estimated" for block values and (right) showing more detail for those blocks containing between 0 and 1000 stones per block

For this and other cut-offs, the percentage of blocks misclassified are as follows:

Cutoff	$10m^2$		$2.16m^2$	
stns/block	%Lost	%Lost Opp.	%Lost	%Lost Opp.
200	1	2	2	2
220	2	2	3	2
240	2	3	3	3
260	2	3	4	3
280	2	3	4	4
300	3	3	6	5
320	4	3	7	5
340	5	4	9	5
360	6	4	8	6
380	7	4	9	6
400	9	5	9	7
Average	3.9	3.3	5.8	4.4

On average, the $10m^2$ samples reduces the misclassification of blocks by around 30% (average total miss-classification of 7.2% vs. 10.2%). The table also shows (on average) a reduced selection of "loss" blocks from 5.8% to 3.9%, a difference of 1.9%.

Furthermore, for the $10m^2$ samples, the lost opportunity blocks reduces by 1.1% which directly contributes to more blocks being available for mining. In total, this relates to approximately 3% more blocks available for mining.

Expressing this in financial terms can only be done approximately because of the varying grades of the "misclassified" blocks relative to the cut-off value and mine plan, but if one approximates that De Beers Marine mines $\pm 5km^2$ per year, this equates for 50x50m blocks with 3% more blocks available, an extra 60 blocks.

At 700 000 carats mined from 2 000 blocks per year at \$240 per carat[5], this implies an improvement in revenue of \$5 million.

If the strategic resource base is quoted for 10 year cycles, this implies a value-add of \$50 million. The figures above are based on a conservative view of the carats mined per year, area mined per year and revenue per carat.

Over and above this direct influence on revenue, the evaluation of the marine deposit in terms of the NPV and strategic valuation of the resource will also improve due to the lowering of the discount rate - as more uncertainty can be removed from the financial risk.

Simulation technology makes this type of improvement possible and research will contribute in a similar manner to other secondary deposits as well.

Other applications of the technology

Similar simulation technology can also be used in exploration and will be applied as soon as the requirements are formulated. In this case, it most probably will not reduce any expenditure, but will optimise the drilling campaigns to get the maximum information from the sampling effort.

[5]These figures are conservative relative to government figures obtained from http://www.namibian.com.na/index.php?id=28&tx_ttnews%5Btt_news%5D=11002&no_cache=1 and http://www.mme.gov.na/gsn/diamond.htm, accessed Setember 2009

B Buffons's needle problem

Buffon's needle problem is a useful method to apply in determining the optimal line spacing in exploration efforts. A description of the method is taken from Van der Watt [65] and summarised below:

B.1 Problem definition

Suppose a large board is ruled with equidistant parallel lines, $2a$ units apart. A thin needle of length $2d < 2a$ is dropped at random on the board. What is the probability that the needle will touch one of the lines?

B.2 Solution

We can think of the ruled board as being made up of a large number of adjacent sections each with a line in the centre and blank strips of width a units on either side of the line. Two adjacent sections are shown below:

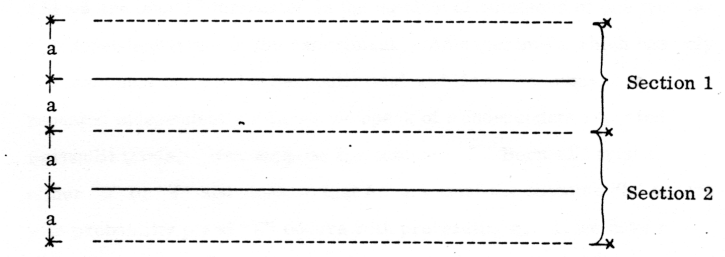
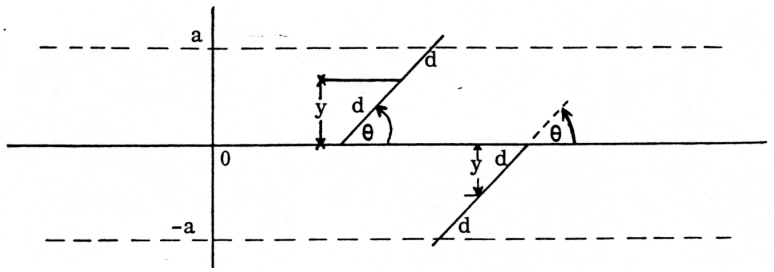


Figure 6.32: Two adjacent sections for Buffons' Problem

We must decide what meaning is to be given to the phrase "dropped at random on the board". We can specify the position of the needle by specifying the coordinates of the centre and the angle it makes width the parallel lines. If we choose the coordinate system in such a way that the x-axis is parallel to the ruled lines, the magnitude of the x coordinate of the centre of the needle will be irrelevant in determining whether or not the needle touches the line. We can therefore suppose that "dropping the needle at random on the board" means that a number is selected at random for the y coordinate of the centre of the needle and an angle θ is chosen at random from the interval $[0, \pi]$, independently of y, for the inclination of the needle to the parallel lines. Because of symmetry, we need only consider what happens if the centre of the needle is assumed to fall in a particular "section". We take the line in this section to be the horizontal axis. The vertical coordinate of the centre of the needle is then to be chosen uniformly random from the interval $[-a, a]$, and the angle of inclination to the horizontal, θ, is chosen at random from the interval $[0, \pi]$.



From the sketch we see that the needle will touch the line in the section (the horizontal axis), if and only if the vertical coordinate of the centre of the needle, y, is chosen in such a way that:

$$-d \ \sin(\theta) \leq y \leq d \ \sin(\theta)$$

for any chosen value of θ in $[0, \pi]$. The event "the needle touches the line" will then be the set of points shown shaded in the diagram below:

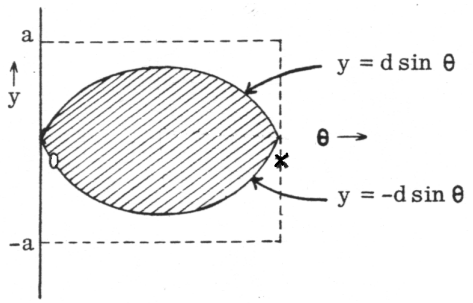


Figure 6.33: Buffons' Problem : Solution area

The area of the shaded region is :

$$2 \int_0^{\pi} d \ \ sin \, \theta \ \ \delta\theta = 4d$$

The area of the rectangle is $2a\pi$ hence:

$$P(\text{needle touches a line}) = \frac{4d}{2a\pi} = \frac{2d}{\pi a}$$

Application of this model is to determine sample line spacing when the mineralisation occurs as long thin lenses with random orientation.

C Sampling challenges in highly dispersed types of mineralisation

W.J. Kleingeld[6], C. Lantuéjoul[7] and C.F. Prins[8]

ABSTRACT

Mineralisation in alluvial diamond or vein gold deposits is often hosted in trapping mechanisms. For instance, diamonds are nested in bedrock depressions such as gullies and potholes which were carved out by the sea long before the diamond deposition took place. Similarly, gold mineralisation can precipitate in the necks of a boudinage structure.

In such deposits, the spatial distribution of the mineralisation is governed by the mutual arrangement of the trapsites and by the grade variations within each trapsite. Reliable estimates can be obtained only if they rest on a proper knowledge of both degrees of variability. This creates a challenging sampling problem, as during the exploration phase it is important to have a good global mean estimate to enable decision making towards the future. A similar problem occurs for local estimates at a later stage of development.

To understand and manage the sampling for global means, specific sampling procedures are required based on an appropriate choice of sample support size and spacing. Estimates from such sampling campaigns should be unbiased, and the less variability they display, the better the sampling parameters. The performance and impact of sampling parameters are tested and compared on different simulated deposits.

INTRODUCTION

One of the most important elements of any mining evaluation exercise is the quality of the information used. Obtaining such information is quite a challenge in environments in which the mineralisation occurs in discrete trapsites (Sichel, 1973) as the samples should pick up both the presence of these traps, as well as the richness (content) thereof.

Typical trapsite systems can be encountered in the Southern African West Coast environment. Such diamond trapsites were formed by wave and/or fluvial processes that eroded potholes and/or gullies into the bedrock. These potholes and gullies formed zones of entrapment for heavier mineral particles including diamonds. Figure 6.34 shows that these trapsites vary in scale from small to very large and occur irrespectively of being on upper beaches (away from the sea), in the surfzone or under the sea. Geologically, the sea level has changed over millions of years and the trapsites have formed before the diamonds were transported to these depositional environments. The presence or size of a trap does not necessarily indicate the richness thereof

[6]Group Manager Mineral Resources, De Beers Group, Mendip Court, Bath Road, Wells, BA5 3DG, United Kingdom. Email: Wynand.Kleingeld@dtc.com

[7]Research Engineer, Ecole des Mines, Centre de Géostatistique, 35 rue Saint-Honoré, 77305 Fontainebleau, France. Email: Christian.Lantuejoul@ensmp.fr

[8]Research Geostatistician, De Beers Group, Mendip Court, Bath Road, Wells, BA5 3DG, United Kingdom. Email: Chris.Prins@dtc.com

and some form of sample measurement is required to determine if an area is financially viable for development.

Figure 6.34: Examples of trapsites (small to large) in an alluvial beach diamond deposit.

In such environments, the area occupied by the trapsites usually does not exceed 10% of the total area of the deposit. Such a low percentage suggests adopting a sampling strategy in two stages, which consists firstly, of detecting the traps and secondly, of assessing the richness of the detected traps. The main advantage of such a strategy is that most of the sampling effort during the second stage is carried out in the vicinity of the detected traps. Accordingly, there is no wasted sampling of the barren regions which could constitute up to 90% of the deposit. Of course, an appropriate estimation procedure has to be set up to compensate for the bias caused by this preferential sampling strategy.

This paper deals mainly with the first stage of this sampling strategy. To define a sampling protocol, one not only has to assign a size, shape and orientation to each individual sample but also to specify the spatial arrangement of the samples. This is investigated in the second part of the paper. The results obtained are illustrated in a third section where it is shown how different primary sampling campaigns affect the variability of the global mean estimate. A discussion follows about the practicality of such sampling campaigns in meeting the required objectives.

DETECTION OF TRAPS

The aim of the primary sampling campaign is not to assess the geometric characteristics nor to evaluate the content of the traps, but only to check whether there exists a trap (or not) at certain locations. The question is how to define an appropriate, if not optimal, sampling protocol.

Suppose that all samples have the same size, shape and orientation as a subset S of $\mathbb{R}^d$. A sample S_x located at x hits the trap T if and only if the point x belongs to a certain subset that depends both on T and S. This subset is called the *dilated set from* T *w.r.t.* S (see Figure 6.35), and usually noted[9] as $T \oplus \check{S}$.

[9]Two notations are introduced here: the point x belongs to the Minkowski sum of A and B, that is

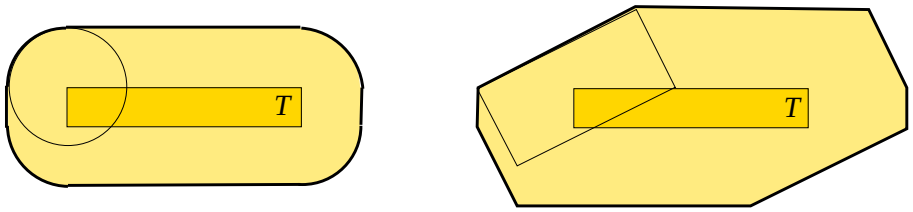


Figure 6.35: Dilation of a rectangular trap w.r.t. a circular (left) or a rectangular (right) sample.

It can be seen that the sample S_x hits the family of traps $(T_i, i \in I)$ if and only if the *point* sample x lies within a trap of the family $(T_i \oplus \check{S}, i \in I)$. This simple remark can be used to study the efficiency of a sampling protocol by *decoupling* the consequences relative to the size, shape and orientation of the samples on one hand, and their spatial location on the other.

Sample size, shape and orientation

The question can be asked: "What is the most appropriate shape for detecting a circular or elongated trap?" For the sake of detection, the sample should be given the best likelihood of hitting the trap. In other words, the sample should be designed so as to maximise the area $a(T \oplus \check{S})$ of $T \oplus \check{S}$.

Of course, such a maximisation procedure cannot be achieved without taking into account physical or financial constraints. For instance, digging very narrow trenches is not technically feasible, and usually the processing cost limits the sample size.

In this paper, the problem of maximising $a(T \oplus \check{S})$ subject to the width constraint $w(S) \geq w_0$ and to the areal constraint $a(S) \leq a_0$ is addressed. Three main cases are discussed:

– If the trap is circular, then the optimal sample is a rectangle with length $\ell_0 = a_0/w_0$ and width w_0;

– If the trap is elongated with orientation θ, then the optimal sample is a rectangle with length ℓ_0, width w_0 and direction θ_0 closely orthogonal to θ (see Figure 6.36).

Figure 6.36: Detection is improved when traps and samples are elongated in orthogonal directions.

$x \in A \oplus B$, iff x can be written as $x = a + b$ with $a \in A$ and $b \in b$. $\check{B}$ is the symmetric of the set B around the origin: $x \in \check{B}$ iff $-x \in B$.

– Occasionally the trap is known to be elongated, but its orientation is unknown. In such a case, it can be shown that rectangular, randomly orientated samples give (on average) better detection prospects than circular samples of the same area.

Proofs of these results are given in the Appendix.

Spatial arrangement of the samples

To get a better understanding of how the spatial arrangement of the samples impacts on the detection of traps, three different sampling configurations were considered (regular, stratified and random). See Figure 6.37.

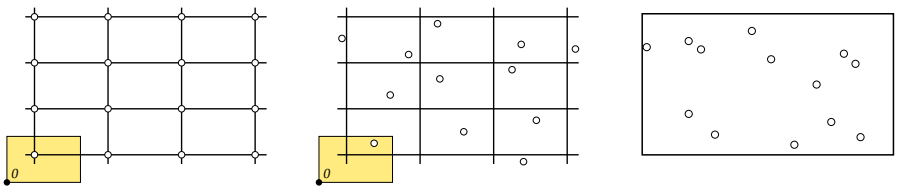


Figure 6.37: Description of three sampling scenarios: regular (left), stratified (middle) and random (right).

Because of decoupling, only point samples are considered, but the sampling intensity (number of samples per unit area) varies. These sampling patterns were applied to two populations of 25 traps, corresponding to two extreme scenarios (regular and random), as depicted in Figure 6.38. Both populations were simulated 10 000 times, and on each simulation the number of traps detected by each sampling pattern was reported.

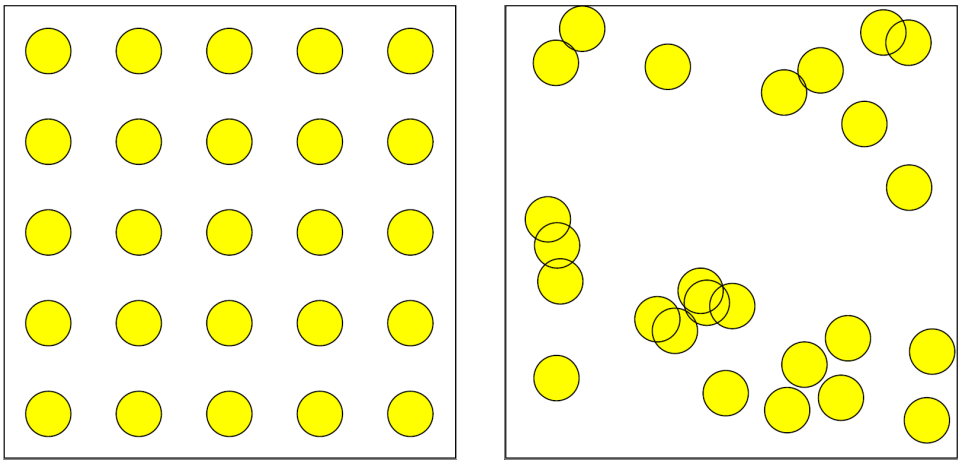


Figure 6.38: Regular (left) and random (right) populations of traps. Both are extreme cases of possible trap locations.

From this exercise, the following conclusions could be drawn:

Firstly, the detection rate appears to be the largest for regularly located samples and turns out to be the smallest for randomly located samples (see Figure 6.39). It should also be pointed out that this figure does not depend on the type of population considered. Indeed, the detection rate is generally affected by the size, shape and orientation of the traps, but not by their intensity nor by their spatial distribution.

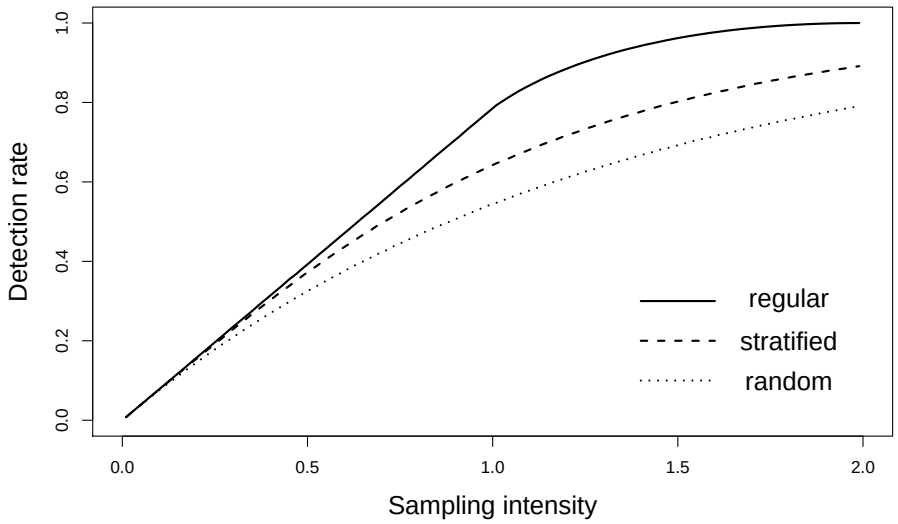


Figure 6.39: Probability of detecting traps using three sampling protocols versus the sampling intensity (number of samples per unit area). They are the same for both populations of traps.

In contrast to this, the histogram of the detected objects depends on the population under study (see Figures 6.40, 6.41 and 6.42). In general, regular sampling yields less dispersed histograms. There are however, some exceptions for regularly located traps, especially when the trapping grid and the sampling grid are multiples of each other (see top left of Figure 6.41). Random sampling seems to produce more dispersed histograms than stratified sampling for random populations.

Finally, regularly spaced samples are more likely to detect traps than any other arrangement of samples (see Figures 6.40, 6.41 and 6.42).

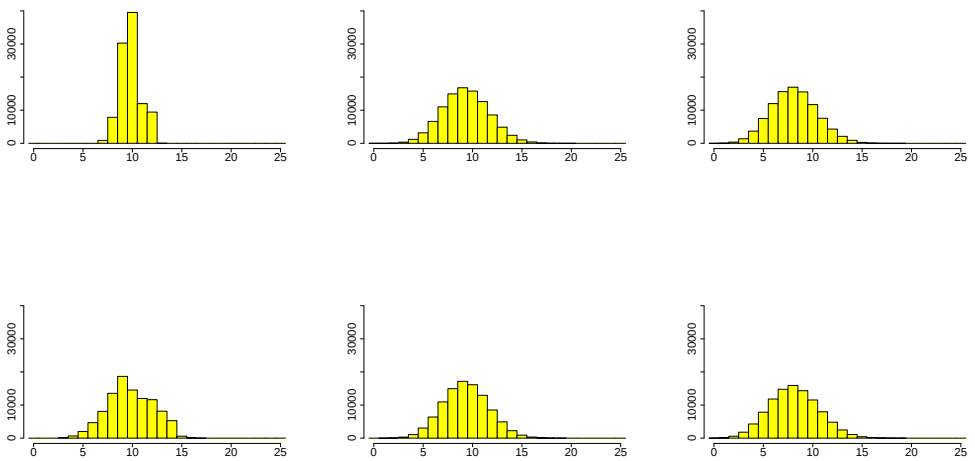


Figure 6.40: Histograms of the numbers of detected traps using a low sampling intensity (0.5). Top and bottom histograms respectively refer to regular and random populations. The impact of the regular (left), stratified (middle) and random (right) sampling scenarios have also been tested.

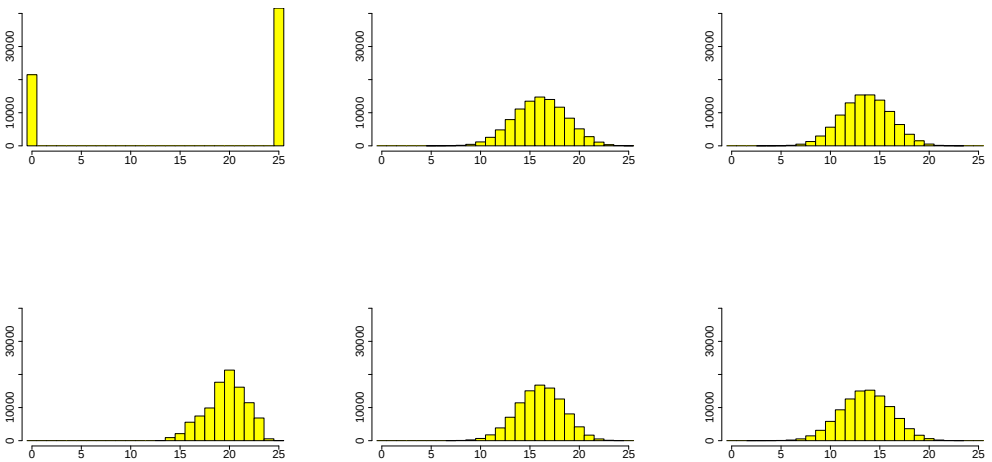


Figure 6.41: Histograms of the numbers of detected traps using an average sampling intensity (1). Top and bottom histograms respectively refer to regular and random populations. The impact of the regular (left), stratified (middle) and random (right) sampling scenarios have also been tested.

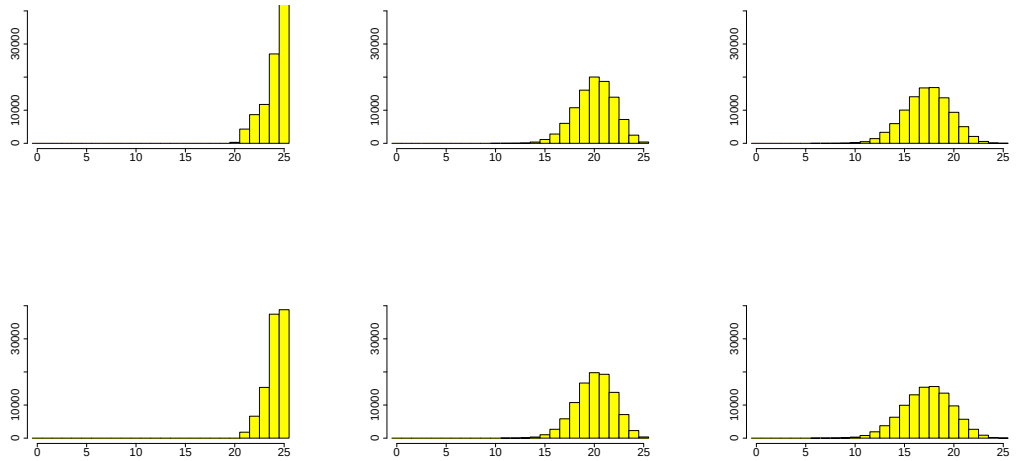


Figure 6.42: Histograms of the numbers of detected traps using a high sampling intensity (1.5). Top and bottom histograms respectively refer to regular and random populations. The impact of the regular (left), stratified (middle) and random (right) sampling scenarios have also been tested.

GLOBAL MEAN ESTIMATION

In some fluvial deposits on the Namibian coast, sample campaigns consist of a small number of samples from which maximum information must be obtained. The ultimate aim from this sampling is to have local estimates in blocks for grade control during the mining stage, but realistically they can often only provide a global mean for a large geological zone.

This section investigates how well primary sampling campaigns perform to estimate the global mean for the number of stones.

To achieve this goal, experiments were made on $100m \times 100m$ simulated deposits. They contain 2 stones per m^2 on average. The traps are circular ($1m$ radius) and randomly located, and the number of stones in each trap is either constant or Poisson. On average, there are 10, 20 or 50 stones per trap, corresponding to the three illustrations of Figure 6.43. For each of the 6 stone number distributions considered (2 types of populations and 3 trapping intensities), 1000 simulations were performed.

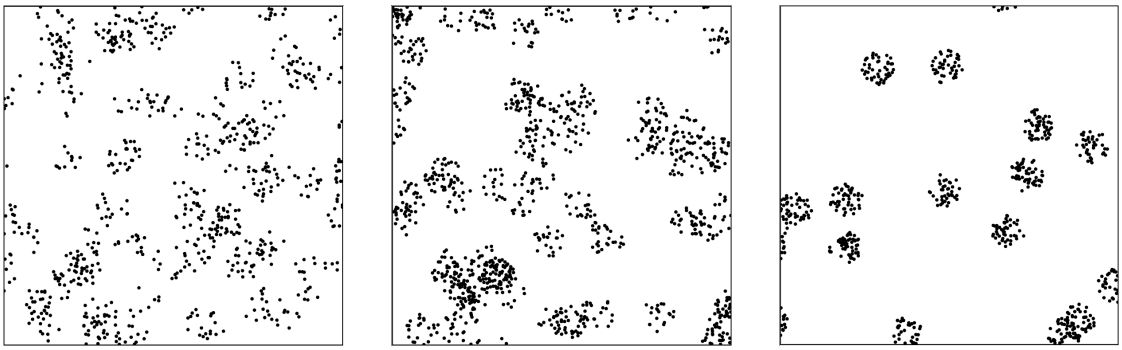


Figure 6.43: Simulations with varying trap and stone densities. In all cases, the intensity has been fixed as 2 stones per m^2.

Regarding the sampling campaigns, the samples are circular and arranged at the nodes of a square grid. All samples of a campaign have the same radius, but different campaigns were used on the same simulated deposit with increasing sample radii ($0.1m$, $0.2m$, $0.4m$, $0.8m$, $1.6m$ and $3.2m$). In all cases, the proportion of sampled area to total area has been fixed as 0.003. The information collected in each sample is the number of stones it contains.

Figure 6.44 shows the cumulative histograms of the estimated number of stones per m^2 for each deposit and each sampling scenario. The corresponding statistics (mean and variance) are presented in Table 1.

Table 1: Mean and variance for the global mean estimates

Model	r = 0.1m	r = 0.2m	r = 0.4m	r = 0.8m	r = 1.6m	r = 3.2m
Constant (10)	2.005	2.000	1.993	2.014	1.992	1.996
	0.067	0.084	0.133	0.296	0.411	0.499
Constant (20)	2.008	1.992	1.966	1.989	2.068	1.933
	0.073	0.102	0.215	0.478	0.834	0.959
Constant (50)	2.001	1.981	1.993	1.976	1.992	2.052
	0.088	0.180	0.436	1.083	2.090	2.693
Poisson (10)	1.993	2.003	1.986	1.972	2.017	2.022
	0.070	0.084	0.138	0.292	0.462	0.559
Poisson (20)	1.999	2.019	1.997	1.996	2.058	2.002
	0.070	0.105	0.204	0.513	0.886	1.126
Poisson (50)	2.007	1.992	2.004	2.023	2.064	2.106
	0.084	0.182	0.449	1.120	2.041	2.644

As expected, no bias is observed on the global mean estimates. It can also be observed that these estimates are more variable as the sample size increases or equivalently as the number of samples decrease. This probably stems from the fact that the range of the simulated models is quite small ($2m$). Accordingly, the samples values are representative only of their immediate vicinity. Note also that even in this rudimentary example, the cumulative histograms show

both sources of fluctuation in the model, namely the trap intensity (number of traps per unit area) and the stone density (number of stones per trap):

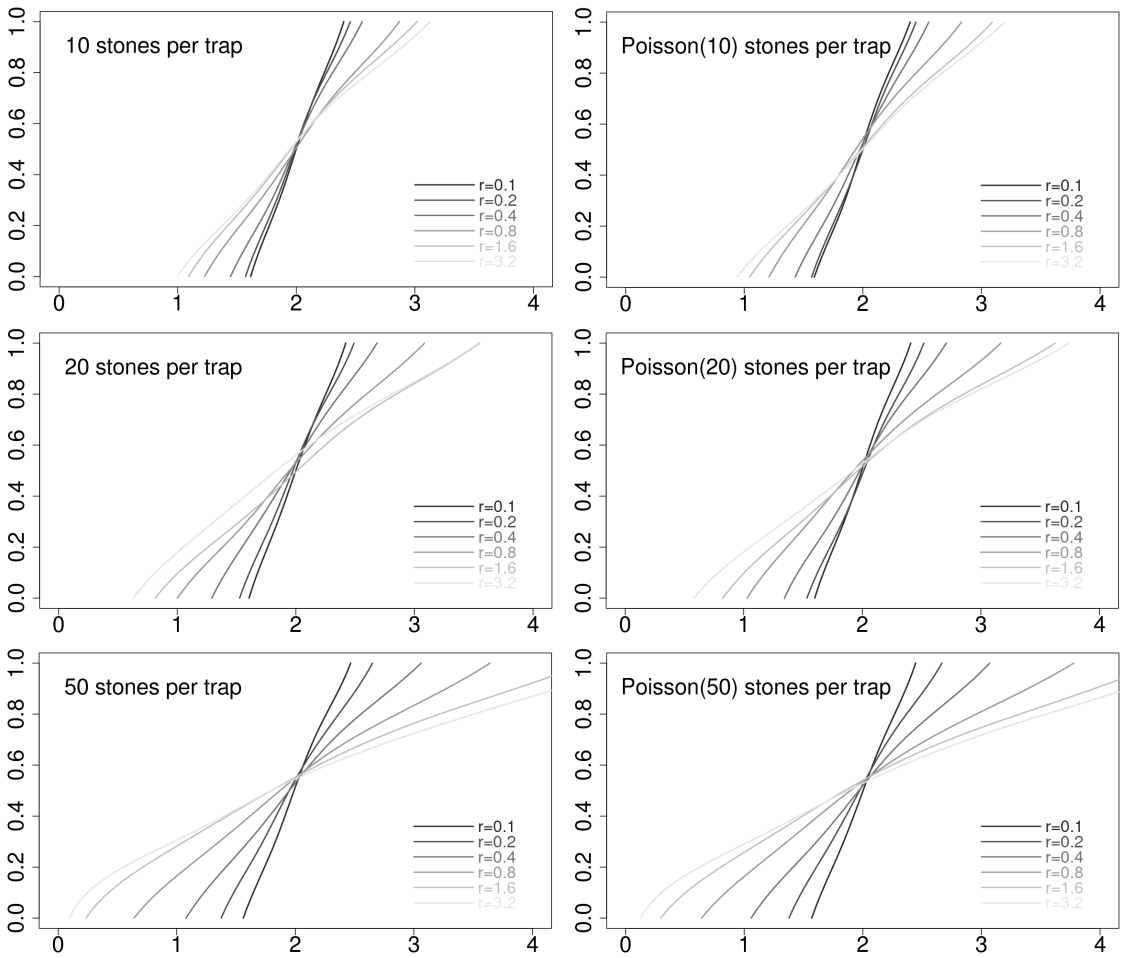


Figure 6.44: Smoothed cumulative distributions for the estimates of the global mean.

– The cumulative histograms of the "pervasive" models (many traps with limited content) are similar for small sample radii ($0.1m$, $0.2m$ and $0.4m$) but differ when the radius becomes large enough ($\geq 0.8m$). This simply expresses the fact that neighbouring samples with small radius are very close to each other. They are redundant and each of them adds very little specific information. This redundancy disappears when the sample radius is large enough. This suggests that pervasive models do not need close sampling to be estimated precisely.

– In contrast to this, the cumulative histograms of the "highly dispersed" models (limited number of traps with high content) have their shape varying continuously as the sample radius increases. In such models, the mineralisation is highly concentrated and accordingly, most small samples miss the traps and are barren, whereas a few of them hit a trap and contain a substantial number of stones. The redundancy exists for small samples, but it is mainly caused

by barren samples. As the sample radius increases, a larger proportion of samples hit traps and the influence of the geometry of the mineralisation gradually disappears (Kleingeld and Lantuéjoul, 1993).

Finally, the cumulative histograms derived from models with constant or Poisson number of stones per trap differ less as the stone density increases. This simply results from the fact that the fluctuations of a Poisson distribution decrease as its parameter increases[10].

DISCUSSION

In this paper, the main question addressed is how a sampling campaign affects the detection of traps. This question which is of prime importance in alluvial diamond mines, can not only be encountered in other mining activities (tailings in coal and sulfide mining, boudinage in gold mining), but also in many other fields. Examples include the detection of pockets of toxic substances left behind in the remediation of an old industrial site, bunches of precious trees in the inventory of a tropical forest, schools of fishes in halieutics, flaws (by non destructive means) in the components of an engine and clumps of dysplastic cells in biological tissue.

Among all features that characterise a sample, shape seems to play a particularly important role. For a given sample area, the more elongated the sample, the more chance it has to hit a trap, and conversely the more circular the sample, the less detection power it possesses. It should be pointed out that many sampling tools are designed to produce circular samples. In this case, it may be interesting to "mimic" an elongated sample by taking a set of aligned, circular samples.

Another observation is that all traps do not have the same chance of being detected. Indeed a sampling campaign with samples S gives the trap T a chance that is proportional to the area $a(T \oplus \check{S})$ of the sample dilated by the trap. In other words, if $F(dS)$ denotes (for short) the distribution (in shape, size...) of the traps, then the distribution of the detected traps is

$$G(dS) \propto F(dS)a(T \oplus \check{S})$$

Since bigger traps have more chance of being detected, the distribution of the detected traps gives the illusion that the traps are bigger in size than they are in reality.

In this paper, particular attention has been paid to the estimation of a global mean. In practice however, local variation in geology requires that sampling must also provide an estimate of the local grade variation which is used for short term mine planning and grade control. This could imply that a different sampling campaign would be required. In evaluating the efficiency of the sampling, a comparison between estimates and actual (see schematic Figure 6.45) is routinely done. The figure shows selected blocks in quadrant 1 which were correctly taken (estimated to be above a cutoff and after extraction provided more than the minimum required grade) and blocks in quadrant 3 correctly left out of the selection process. What is important are the blocks in quadrant 2, which will result in financial loss and blocks in quadrant 4, which were not mined and represents lost opportunity. It is strategically extremely important when dealing with a single mean estimate for an area to have the estimate within the correct quadrant. The

[10]The coefficient of variation of a Poisson distribution with parameter θ is $1/\sqrt{\theta}$.

same also applies later during the development of the project for the smaller selected mining blocks. The second ellipse (not at 45° angle) shows the catastrophic effect that bias can cause.

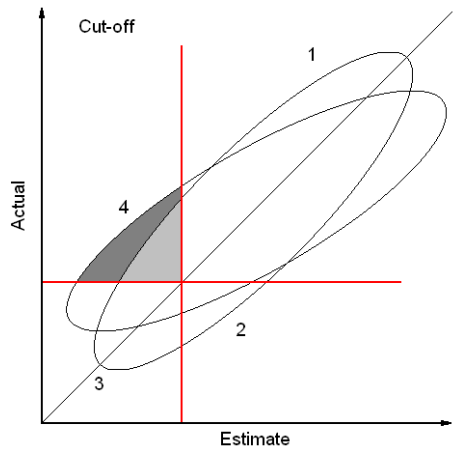


Figure 6.45: Comparison between actual and estimated grades.

Usually an intermediary sampling campaign is required that can give information on both a strategic level (deciding on the continuation during the exploration phase - as a first decision), as well as providing information on the right scale for local estimates. Although more smaller samples would be the right strategy for a global estimate, certain practical problems are created when the samples become too small. For example, the smaller samples usually lack precision and representativity problems in content. This in turn and also due to the low grades usually expected, will create more samples with zero values, causing a loss of kriging efficiency due to the departure from Gaussian assumptions.

If a strategy for such a situation must be determined, emphasis must be placed on and a clear objective stated before embarking on the sampling campaign. A sample campaign optimised for one objective might not be appropriate for the other objective.

CONCLUSION

A global mean estimate is the first value determined during the exploration phase to show whether the sampling should continue to the pre-feasibility stage or not. For this reason it is important to design a sampling campaign that reduces the uncertainty of such an estimate as much as possible.

In this paper, various sampling campaigns have been tested to estimate a global mean in a geological environment with trapping mechanisms. From these experiments it appears, firstly that rectangular samples are better than circular samples, secondly, that it is better to have the rectangular samples orientated in random directions if no prior geological knowledge exists about the trapsite direction of occurrence, thirdly, that more smaller samples are better than less, but larger samples and lastly that a regular grid is superior to randomly located samples.

Of course, the eventual estimation of local estimates must be kept in mind when the sample parameters are determined. In this highly dispersed mineralisation environments, further research is therefore required into the sampling challenges and estimation of the in-situ material.

Acknowledgement: The authors want to thank the reviewer for his constructive criticism of the paper, and also for suggesting several domains of application where the proposed method could be applied.

REFERENCES

Hadwiger, H. 1957. *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie*, 312 p (Springer-Verlag: Berlin)

Kleingeld, W.J. and Lantuéjoul, C. 1993. Some comments on the sampling of highly dispersed orebodies, in *Geostatistics Tróia'92* (Ed: A. Soares), pp. 953-964 (Kluwer: Dordrecht)

Sichel, H.S. 1973. Statistical valuation of diamondiferous deposits, in *Application of computer methods in the mineral industry* (Eds: M.G.D. Salamon and F.H. Lancaster), pp. 17-25 (The South African Institute of Mining and Metallurgy: Johannesburg)

APPENDIX: SIZE, SHAPE AND ORIENTATION OF SAMPLES

There is no practical inconvenience in assuming that traps and samples are convex.

Suppose that the trap T can be written as $T = L \oplus D(r)$ where L is a segment with length b and $D(r)$ is a disk of radius r. Consider a rectangular sample S with length ℓ, width w and orientation θ w.r.t. L ($0 \leq \theta < \pi$). The area and the perimeter of $L \oplus \check{S}$ are respectively given by

$$a(L \oplus \check{S}) = b\ell \sin\theta + bw|\cos\theta| + \ell w \qquad p(L \oplus \check{S}) = 2(b + \ell + w)$$

(see Figure A).

Figure A

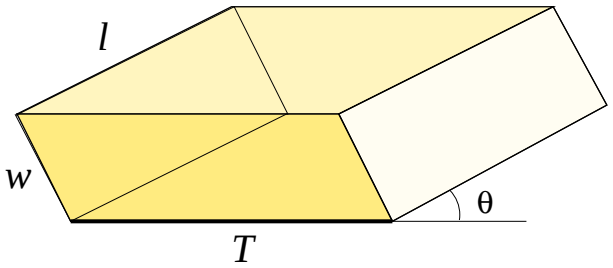


Figure A: Dilation of a segment by a rectangle

Consequently, and owing to Steiner's formula

$$a\big(X \oplus D(r)\big) = a(X) + rp(X) + \pi r^2$$

if X is convex (Hadwiger, 1957), the area of $T \oplus \check{S}$ is

$$\begin{aligned} a(T \oplus \check{S}) &= a\big((L \oplus \check{S}) \oplus D(r)\big) = a(L \oplus \check{S}) + rp(L \oplus \check{S}) + \pi r^2 \\ &= b\ell \sin\theta + bw|\cos\theta| + \ell w + 2r(b + \ell + w) + \pi r^2 \end{aligned}$$

The problem is to maximise this area when $\ell w = a_0$, $\ell \geq w \geq w_0$ and $0 \leq \theta < \pi$.

Because of the constraint $\ell w = a_0$, and based on symmetry arguments, the problem boils down to maximising the function

$$F(w, \theta) = b\frac{a_0}{w}\sin\theta + bw\cos\theta + a_0 + 2r\left(b + \frac{a_0}{w} + w\right) + \pi r^2 \tag{6.1}$$

subject to the constraints $w_0 \leq w \leq \sqrt{a_0}$ and $0 \leq \theta \leq \pi/2$.

Sample optimisation

Suppose at first that w has been fixed. The θ value that maximises F satisfies

$$\frac{\partial F}{\partial \theta} = b\frac{a_0}{w}\cos\theta - bw\sin\theta = 0$$

that is

$$\theta = \arctan\frac{a_0}{w^2}$$

Substituting this value into F, one obtains after some simplifications

$$F(w) = \frac{b}{w}\sqrt{a_0^2 + w^4} + a_0 + 2r\left(b + \frac{a_0}{w} + w\right) + \pi r^2$$

By differentiation, it is then easy to establish that F has a single minimum at $\sqrt{a_0}$. It follows that the maximum of F on $[w_0, \sqrt{a_0}]$ is precisely at w_0. The corresponding optimal direction θ_0 is

$$\theta_0 = \arctan\frac{a_0}{w_0^2}$$

This direction θ_0 is close to $\pi/2$ when w_0 is small.

Comparison between circular and rectangular samples

It is interesting to compare the optimal dilated area obtained

$$F(w_0) = \frac{b}{w_0}\sqrt{a_0^2 + w_0^4} + a_0 + 2r\left(b + \frac{a_0}{w_0} + w_0\right) + \pi r^2$$

with a circular sample with area $a_0 = \pi r_0^2$:

$$G(r_0) = 2b(r + r_0) + \pi(r + r_0)^2 = 2b\left(r + \sqrt{\frac{a_0}{\pi}}\right) + \pi\left(r + \sqrt{\frac{a_0}{\pi}}\right)^2 \tag{6.2}$$

Let $\lambda = a_0/w_0^2$ and $\mu = b/r$. One obtains by difference

$$\begin{aligned}\frac{1}{w_0 b}\left(F(w_0) - G(r_0)\right) &= \mu\sqrt{\lambda^2+1} + 2(\lambda+1) - 2\sqrt{\frac{\lambda}{\pi}}(\mu+\pi) \\ &= \mu\left(\sqrt{\lambda^2+1} - 2\sqrt{\frac{\lambda}{\pi}}\right) + 2(\lambda+1) - 2\sqrt{\lambda\pi}\end{aligned}$$

Now, it can also be seen that both functions $\lambda \longrightarrow \sqrt{\lambda^2+1} - 2\sqrt{\lambda/\pi}$ and $\lambda \longrightarrow \lambda + 1 - \sqrt{\lambda\pi}$ are positive and increasing on $[1, \infty[$. This is enough to conclude that $F(w_0) > G(r_0)$. In other words, a rectangular sample performs better than a circular one.

It should be pointed out that the obtained result is valid for each $b > 0$. Letting b tend to 0, one obtains that circular traps are better detected by rectangular samples than by circular ones.

Case where the trap orientation is unknown

In this case, there is no reason to assign a preferential direction to the sample. Accordingly, assume its orientation uniform on $[0, \pi[$. Because of (6.1), the expected dilated area can be written as

$$E\{a(T \oplus \check{S})\} = \frac{2b}{\pi}\left(\frac{a_0}{w} + w\right) + a_0 + 2r\left(b + \frac{a_0}{w} + w\right) + \pi r^2$$

It turns out that this dilated area is always larger than the one obtained from a circular sample with area a_0. Indeed, subtracting equation (6.2) from the previous equation, one obtains after successive simplifications

$$E\{a(T \oplus \check{S})\} - G(r_0) = 2\left(\frac{b}{\pi} + r\right)\left(\frac{a_0}{w} + w - \sqrt{\pi a_0}\right)$$

The right hand side member of this equation is positive whatever the value of w.

* * * * * * * * * * *

(Deliberately left blank)

D Kriging using directional neighbourhoods

D.1 Introduction

New areas for exploration have been identified and considered for mining in the marine environment off the west coast of Namibia. Marine gravel, diamonds and sand have been deposited in marine channels that acted as entrapment features. Some parts of the channels are exposed and visible, but others are blanketed by sand and not visible on side scan sonar images. In Figure B the darker shaded areas were digitised and represent bedrock outcrop and the lighter coloured areas could either not be interpreted because of the blanketing or are expected to be channels containing gravels and sand.

There is a prominent directional character to the entrapment features which are related to the preferential weathering along lineations in the footwall. These features are not equidirectional, which poses a challenge when anisotropy is considered when kriging or simulation is carried out.

If the assumption is made that the mineralisation is related to these features, an approach is needed to model and incorporate this information into the estimation process. The concept of directional kriging, which was formulated (but not applied) during a project in 1996 for Hudson Bay Mining and Smelting by Prins [55], is relevant. The idea was developed to cater for the estimation of minerals in environments with very complex folding that could not be done in other ways.

There are similarities between the directional aspect of the Hudson Bay deposits and the gullied marine environment. Similar to other publications by Boisvert [6, 7, 8] and Isaaks [34], a method making use of a local directional component (neighbourhood and/or anisotropy) for the kriging process is proposed. For the following arguments, assume that Figure B describes the target area to be estimated.

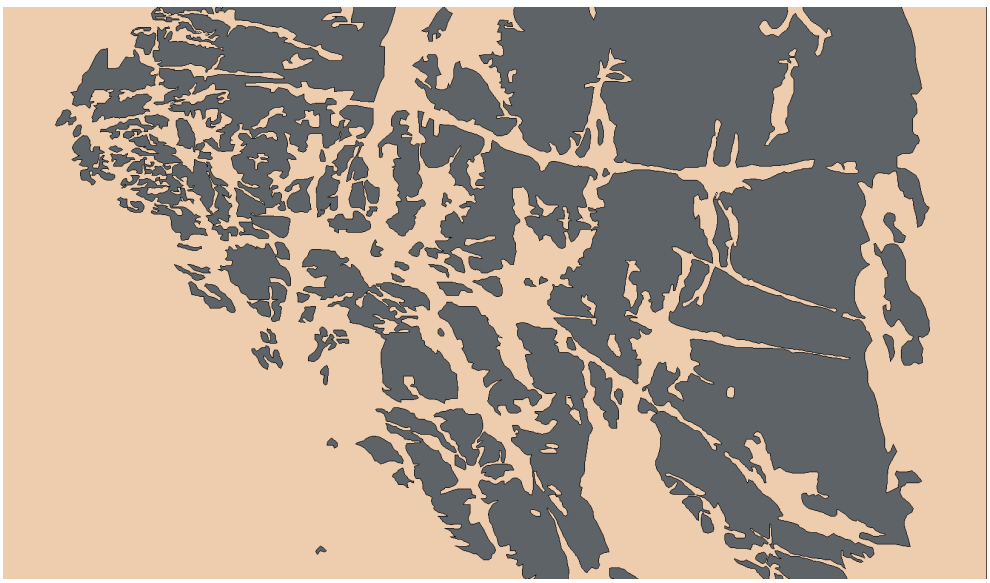


Figure B: Directional character of the underlying entrapment features

Through interpretation of side scan sonar images, seismic data and regional geological interpretation, the directional component can be modelled and incorporated into the estimation process. In this way, the accuracy of estimation should improve.

For example, Figure B above can be interpreted as:

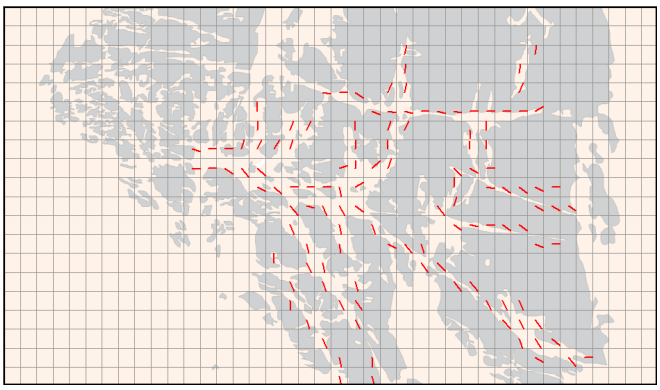


Figure C: Anisotropy interpretation

Sampling and mining these environments raises a number of challenges, depending on the scale and size of the channelled features. Very accurate placement of the sample tool and mining equipment would be a prerequisite and is assumed to be possible.

Directional interpretation for the grey blocks in Figure C is not important as these blocks will not be mined.

Visualisation of geological features is possible using the latest remote sensing methods and interpretation methods and visualisation equipment have improved dramatically. Swath bathymetry and side scan sonar examples are shown.

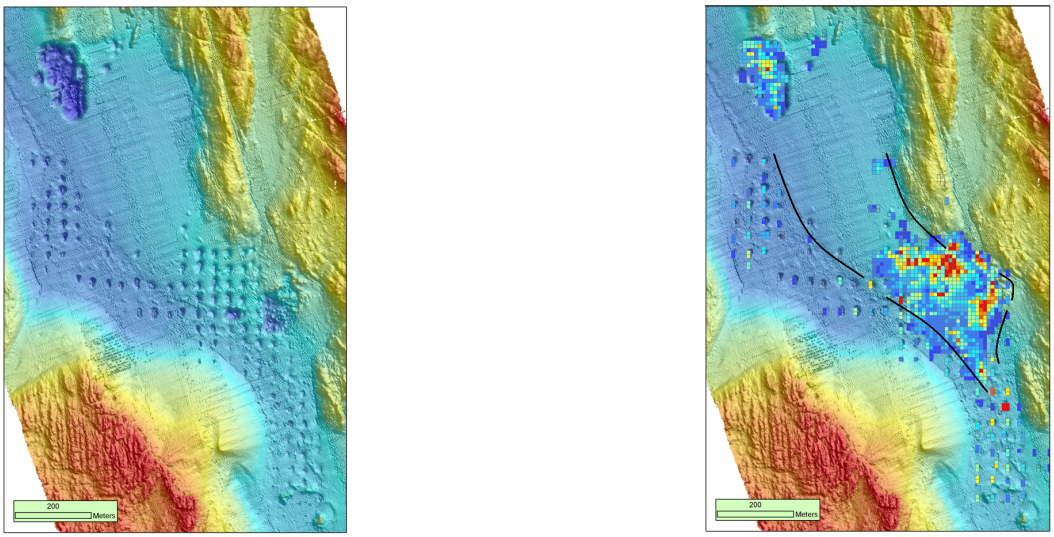


Figure D: Swath Bathymetry images showing topography, bedrock outcrop and sample holes. Some test mining is shown with regional anisotropy.

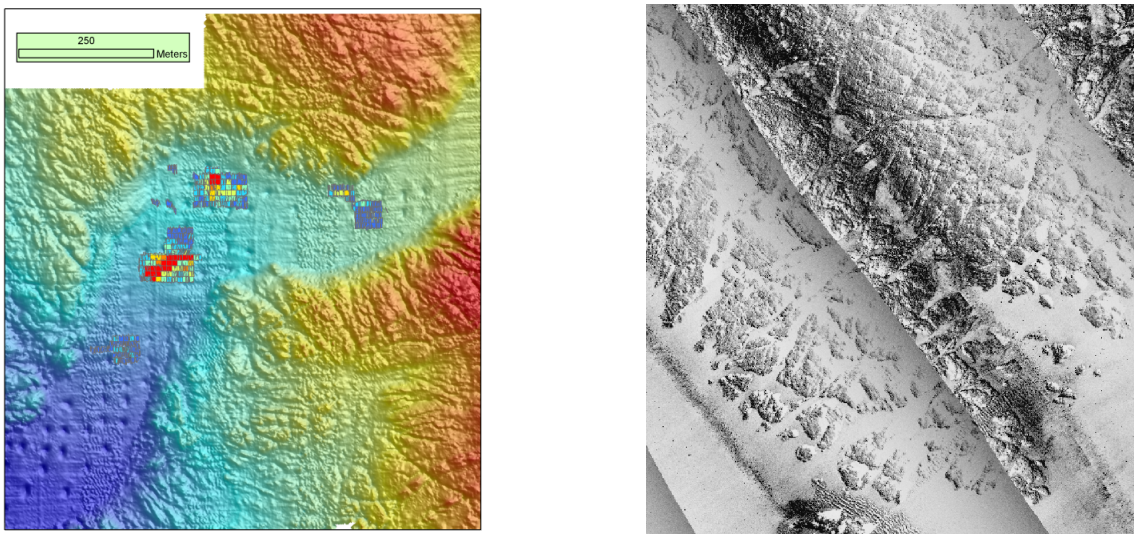


Figure E: Swath Bathymetry (left): Geological features, sample holes and some test mining. (Right): Side scan sonar image showing channels (right)

The images in Figures D and E show the channels, surrounding footwall, samples and test mining. With the local direction modelled and parameterised, the kriging procedure is performed as usual.

The behaviour of kriging weights using directional neighbourhoods warrant further comment.

The objective is to krige the centre block of a 5x5 kriging plan with typical variograms observed in the marine environment and to summarise the behaviour of the kriging weights, depending on the sample configuration. The neighbourhoods are selected as an isotropic and anisotropic neighbourhood, with the anisotropic direction along the 135^o direction (as often observed in marine environments). The anisotropy and directional sample selection is done is the same direction.

D.2 Kriging weights using directional neighbourhoods

Software packages make provision for modelling anisotropy in the variography and the search neighbourhood used in the kriging. With the incorporation of a varying directional component to cater for gullied environments, the kriging neighbourhood and variogram can be applied in different combinations:

a) Using a local anisotropic variogram and isotropic neighbourhood The neighbourhood's samples are not only selected along the channel direction, and only the anisotropy of the variogram will effect the kriging weights

b) Using a local anisotropic variogram and local anisotropic neighbourhood The neighbourhood and variogram are both aligned with the direction of the channel feature and kriging weights are determined.

Tests were conducted using combinations of variogram and neighbourhood parameters on a 5x5 kriging plan for two typical extremes in variogram models. Double structured variograms

were used and the nugget effect to sill ratio was varied for the two models to mimic typical spatial structures observed in the marine environment.

Variograms with nugget effect equal to 40% of total sill:

Isotropic variogram (40% nugget effect):

$$\gamma(\text{h}) = 0.4 + \text{Spherical}\ (Sill_1 = 0.3,\ Range_1 = 100\text{m},$$
$$+\ \text{Spherical}\ (Sill_2 = 0.3,\ Range_2 = 280\text{m})$$

The anisotropic variogram was given a range of 2/3 of the isotropic range and direction 45^o (40% nugget effect):

$$\gamma(\text{h}) = 0.4 + \text{Spherical}\ (Sill_1 = 0.3,\ Range_1 = 100\text{m \& } 66\text{m}),$$
$$+\ \text{Spherical}\ (Sill_2 = 0.3,\ Range_2 = 280\text{m \& } 186\text{m}).$$

Variograms with high nugget effect equal to 85% of the total sill:

Isotropic variogram (85% nugget effect):

$$\gamma(\text{h}) = 0.85 + \text{Spherical}\ (Sill_1 = 0.075,\ Range_1 = 100\text{m},$$
$$+\ \text{Spherical}\ (Sill_2 = 0.075,\ Range_2 = 280\text{m})$$

The anisotropic variogram was given a range of 2/3 of the isotropic range and direction 45^o (85% nugget effect):

$$\gamma(\text{h}) = 0.85 + \text{Spherical}\ (Sill_1 = 0.075,\ Range_1 = 100\text{m \& } 66\text{m}),$$
$$+\ \text{Spherical}\ (Sill_2 = 0.075,\ Range_2 = 280\text{m \& } 186\text{m}).$$

The test results when kriging regularly spaced data with different variogram configurations are shown in Figures F and G.

D.3 Interpretation of results

The weights for individual blocks have been grouped to express the difference in weight distribution for blocks along the direction of the channel (main direction of anisotropy), versus the blocks across the channel (perpendicular to the main anisotropy direction).

Schematically:

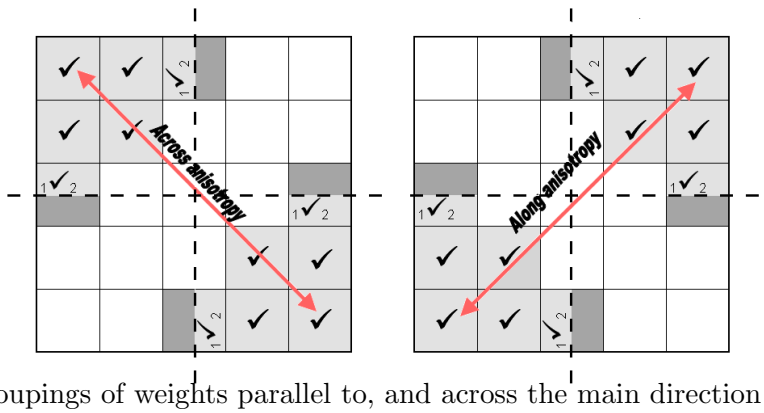


Figure H: Groupings of weights parallel to, and across the main direction of anisotropy

The two sets of weights, across and along the anisotropy direction in Figure H, have the split

Isotropic variogram: Low nugget effect

0.8%	0.6%	0.7%	0.6%	0.8%
0.6%	3.9%	10.6%	3.9%	0.6%
0.7%	10.6%	31.0%	10.6%	0.7%
0.6%	3.9%	10.6%	3.9%	0.6%
0.8%	0.6%	0.7%	0.6%	0.8%

Isotropic variogram : High nugget effect

2.8%	3.0%	3.2%	3.0%	2.8%
3.0%	4.3%	6.1%	4.3%	3.0%
3.2%	6.1%	10.4%	6.1%	3.2%
3.0%	4.3%	6.1%	4.3%	3.0%
2.8%	3.0%	3.2%	3.0%	2.8%

Anisotropic variogram: Low nugget effect

0.7%	0.2%	0.5%	0.8%	0.2%
0.2%	2.3%	10.7%	8.2%	0.8%
0.5%	10.7%	28.2%	10.7%	0.5%
0.8%	8.2%	10.7%	2.3%	0.2%
0.2%	0.8%	0.5%	0.2%	0.7%

Anisotropic variogram: High nugget effect

2.5%	2.6%	3.1%	3.4%	3.0%
2.6%	3.7%	6.1%	5.5%	3.4%
3.1%	6.1%	9.8%	6.1%	3.1%
3.4%	5.5%	6.1%	3.7%	2.6%
3.0%	3.4%	3.1%	2.6%	2.5%

Figure F: Kriging weights for a 5x5 isotropic search neighbourhood

Isotropic variogram: Low nugget effect

		1.0%	0.8%	1.0%
	4.4%	10.7%	4.0%	0.8%
1.0%	10.7%	31.1%	10.7%	1.0%
0.8%	4.0%	10.7%	4.4%	
1.0%	0.8%	1.0%		

Isotropic variogram: High nugget effect

		4.2%	3.8%	3.7%
	5.5%	7.0%	5.1%	3.8%
4.2%	7.0%	11.2%	7.0%	4.2%
3.8%	5.1%	7.0%	5.5%	
3.7%	3.8%	4.2%		

Anisotropic variogram: Low nugget effect

		0.9%	0.3%	0.9%
	8.7%	10.8%	2.3%	0.3%
0.9%	10.8%	28.2%	10.8%	0.9%
0.3%	2.3%	10.8%	8.7%	
0.9%	0.3%	0.9%		

Anisotropic variogram: High nugget effect

		4.2%	3.5%	3.5%
	6.9%	7.1%	4.6%	3.5%
4.2%	7.1%	10.7%	7.1%	4.2%
3.5%	4.6%	7.1%	6.9%	
3.5%	3.5%	4.2%		

Figure G: Kriging weights for an anisotropic search neighbourhood (elliptical search neighbourhood)

blocks in common for which the weight was split evenly between the two groups. The blocks immediately surrounding the centre block have symmetrical weights and are not included in the percentage expressions of weight. The groupings express the percentage weight for comparison along and across the direction of anisotropy.

In the following figures, each informed block contains a dot and the sum of the percentage of kriging weight along and across the main direction of anisotropy are shown as a percentage.

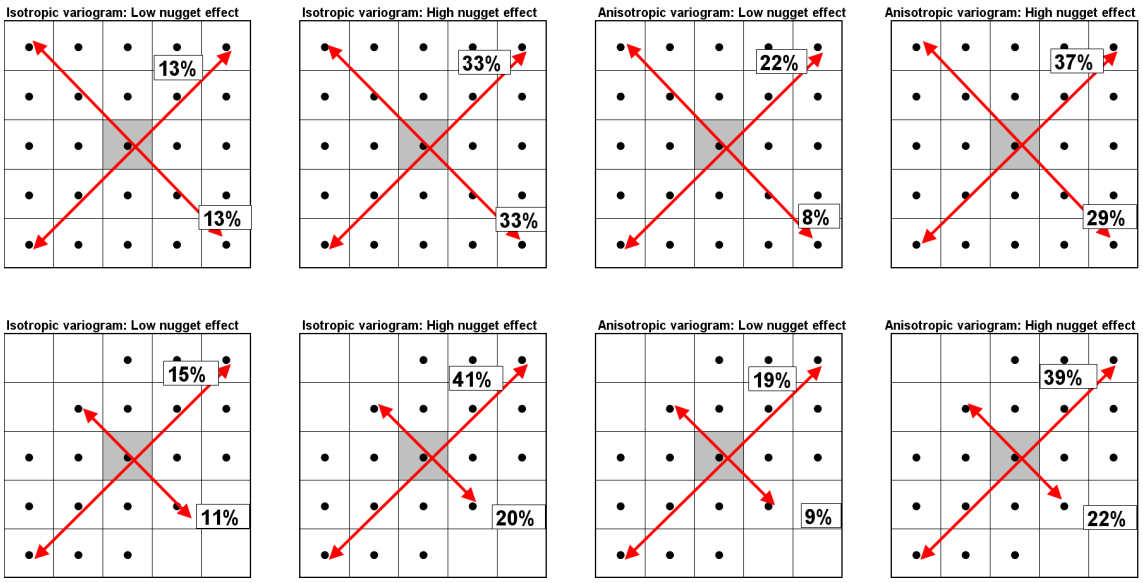


Figure I: Groupings of blocks across and parallel to the direction of anisotropy. Top line with isotropic and bottom line with anisotropic search neighbourhoods (direction 45^o)

Comments:

For isotropic 5x5 neighbourhood (top row in Figure I):

• For isotropic variograms and neighbourhoods, as expected, the weights with and across the neighbourhood are evenly distributed for variograms with low and high nugget effect ratios.

• With anisotropy applied only to the variogram (and not the selection of samples for the neighbourhood), the low nugget effect case assigns, as expected, more weight along the direction of anisotropy than across (22% vs. 8%). The same hold for the high nugget effect but to a lesser extent (37% vs. 29%).

For anisotropic neighbourhoods (bottom row in Figure I):

• Both the isotropic and anisotropic variograms, combined with anisotropic neighbourhoods, show results that assign additional weighting to the samples along the main axis of the neighbourhood's direction.

The selection of samples within a neighbourhood aligned with the anisotropy direction causes the weight distribution to be surprisingly close to the results from the use of the variograms with anisotropy.

This warrants some more comment:

If the results using the anisotropic variograms for the two types of neighbourhoods are considered, the weighing is approximately the same (22% and 8% versus 19% and 9%, also 37% and 29% versus 39% and 22%). In practice though, the kriged blocks for the anisotropic neighbourhoods will be constrained much more along the anisotropy direction due to the effect of sample selection. Kriging along channels and selecting the neighbourhood accordingly will emphasise and align the kriged results more strongly with the channel than for an isotropic neighbourhood.

D.4 Final Remark

The examples and results of the study cannot be proposed as a generalised result as it is case specific. Each sample, block and neighbourhood configuration must be tested when directional kriging is used to ensure that the approach is fit for purpose.

(Deliberately left blank)

E Summary of application of Gy's sampling theory

P. Gy's [33] research on sampling theory and subsequent guidance to quantify and implement good sampling practices through the application of Gy's formula, identifies different components contributing to the overall "error" made in the sampling and estimation process. Acknowledging and managing the different components can dramatically improve the quality of sampling and derived estimates. A summary of the components comprising the total estimation error is quoted from François-Bongarçon [25]:

TOTAL ESTIMATION ERROR: TE = SE + GE + IE

- o **TOTAL SAMPLING ERROR:** SE = CE + AE
  - - ***SAMPLE COLLECTION ERROR*:** CE = LE + ME + FE
    - LOCALIZATION ERROR: LE
    - MATERIALIZATION ERROR: ME = DE + TE
      Sample delimitation and extraction error: DE
      Sample taking error (grouping/segregation): TE
    - FUNDAMENTAL ERROR: FE
  - - ***SAMPLE ANALYTICAL ERROR*:** AE
- o **GEOSTATISTICAL ESTIMATION ERROR:** GE = ME + KE
  - - ***MODELIZATION ERROR*:** ME
  - - ***KRIGING/ESTIMATION ERROR*:** KE
- o **GEOLOGICAL INTERPRETATION ERROR:** IE

The error describing how the sample size effects the representivity of samples is the fundamental error (FE), represented as σ^2_{FE}.

E.1 Description of factors and constants in Gy's fundamental error

The formulation of the fundamental error as derived by P. Gy [33] and interpreted by Royle[59] is expressed as:

$$\sigma^2_{FE} = \left(\frac{1}{M_S} - \frac{1}{M_L}\right) c\ f\ l\ g\ d^3\ A^2$$

The parameters and constants are as follows:

M_S **:** This is the mass of the sample (grams)

M_L **:** This is the mass of the lot (grams)

c **:** The mineralogical composition factor. This takes into account the volumetric mass and proportion of constituent elements. In the case of material with 2 components:

$$c = \frac{(1-a_L)}{a_L}\left[(1-a_L)D_A + a_L D_G\right]$$

where

a_L : decimal proportion of mineral A in L (as a proportion). Ex: 1% becomes 0.01

D_A : Mean volumetric mass of mineral (diamond = 3.52g/cm^3)

D_G : Mean volumetric mass of all components excluding mineral A i.e. gangue

For rare materials, c simplifies to

$$c = \frac{D_A}{a_L}$$

l : is the liberation factor with $0 \leq l \leq 1$.

$l = 0$: If all components constitute perfectly homogenous mixture. This is non-realistic and hardly ever observed.

$l = 1$: Take $l = 1$ where all components are perfectly liberated. Due to the particulate nature of diamonds and for purposes of alluvial/dump deposits where diamonds will mostly be liberated in the mining process, it is sufficient to assume that $l = 1$.

f : is the parameter that describes the dimensionless shape of a fragment. This value is derived from the fraction of the actual volume of the fragment to that of a cube with the same length as the sieve aperture that allows the fragment to pass through it. For example, take a cube of size $d \times d \times d$ and accompanying sphere with radius r, just fitting through the sieve with $d = 2r$. The ratio of the volume of the sphere $\left(\text{Vol}_{sphere} = \frac{4}{3}\pi r^3\right)$ and the accompanying cube with volume $\left(\text{Vol}_{cube} = d^3 = 8r^3\right)$ is 0.52.

For a spherical shape assign $f = 0.52$ and for most compact fragments a value close to $f = 0.5$ must be used.

g : is the parameter that shows how tight the granulometric distribution of the fragments are. Gy gives typical values as:

$g = 0.25$: Natural un-calibrated populations

$g = 0.40$: Populations calibrated from above. For example populations passing through a sieve. There is a maximum truncated size.

$g = 0.50$: Oversize and calibrated from below

$g = 0.6/0.75$: Calibrated from both above and below

$g = 0.75$: Populations that are naturally calibrated (Example : carob seeds)

$g = 1.00$: Populations that are perfectly calibrated. Example : Balls with the same radius

d : The d_{95} value is the size through which 95% of the material fits in terms of it size (expressed in cm).

A : The grade of the mineral expressed in g/m^3

Knowing the physical characteristics of the material to be sampled and some idea of the expected grade (or minimum required threshold grade), σ^2_{FE} can be used to determine a sample size and express the expected confidence limits when samples are collected.

F The conditional simulation of a Cox process with application to deposits with discrete particles

The conditional Cox process is recognised as the simulation method of choice for the simulation of diamond deposits and is extensively used in sample optimisation studies in secondary diamond deposits. A summary from the paper by Kleingeld, et al. [43] is given.

F.1 The Cox process

With the mineralisation as discrete particles, the model considered is the Cox process, after work done by Kleingeld and Lantuéjoul [39, 44].

In secondary diamond deposits, the distribution of the particles is affected by factors such as the particle source, local geography and footwall lithology, without the concentration (or absence) being contributed to a single factor. The combined influence of all the factors is grouped into the concept of potential. This potential represents the propensity of richness of mineralisation for the region to be simulated.

For use in a statistical model, the potential is considered as a regionalised variable with some areas richer in potential than others.

Sichel [60] suggests the idea of modelling the number of stones in two steps, firstly, to model the sample potential and secondly, the number of stones once the potential has been fixed. The Cox process developed is the spatial transcription of the method elaborated by Sichel.

Quoting from the paper by Kleingeld, et al.: "In a formal manner, the Cox process can be described as an implantation of discrete particles according to a Poisson point process for which the regionalised intensity is a random function."

This Cox process can be characterised by it's spatial law, the total of all values taken by the probabilities

$$P\{\bigwedge_{i\in I} N(v_i) = n_i\} \;=\; E\;\{\prod_{i\in I} exp\{-Z(v_i)\;\}\;\frac{Z^{n_i}(v_i)}{n_i!}\} \tag{6.3}$$

From (6.3) the spatial law of the Cox process is entirely specified by the spatial law of the potential.

Suppose the deposit is known by a number of samples equal in size and shape with available distribution of the number of particles and semi-variogram. If the potential at point x is designated by Z_x and the number of particles by N_x, where $Z_x = Z(v_x)$ and $N_x = N(v_x)$, then

$$P\{N_x = n\} = E\{exp(-Z_x)\frac{Z_x^n}{n!}\} \tag{6.4}$$

There is an inversion formula for 6.4 which permits to deduce the law of Z_x from that of N_x, but for the application it suffices to refer to known results. For example if N_x follow a negative binomial distribution then it can be shown that Z_x follow a gamma distribution (Feller,[23]) and similarly when N_x follows a Sichel distribution, then Z_x follows an inverse Gaussian distribution (Sichel, [60]).

Concerning the covariance of the potential

$$Cov\{Z_x, Z_{x+h}\} = Cov\{N_x, N_{x+h}\} - E\{N(v_x \cap v_{x+h})\} \tag{6.5}$$

Thus, the covariance function of the potential coincides with that of the particle count, with the exception of the term $E\{N(v_x \cap v_{x+h})\}$, which results in the Poisson implantation of the particles. This is zero if the two terms v_x and v_{x+h} are disjoint and is empirically seen as a nugget effect.

Knowing from the experimental information the distribution and covariance of Z_x is not sufficient to define the spatial law of Z unless Z is an anamorphosed multigaussian random function. Z is made an anamorphosed multigaussian function where Y is a standard stationary multigaussian random function by:

$$Z_x = \varphi(Y_x) \tag{6.6}$$

If φ is expanded in terms of Hermite polynomials we have

$$\varphi(y) = \sum_{n=0}^{+\infty} \frac{\varphi_n}{n!} H_n(y) \tag{6.7}$$

In addition one knows that the covariance of the potential satisfies

$$Cov\{Z_x, Z_{x+h}\} = \sum_{n=1}^{+\infty} \frac{\varphi_n^2}{n!} \rho^n(h) \tag{6.8}$$

where ρ is the correlation function of Y.

Knowing the anamorphosis which can be derived from the distribution of Z_x, the covariance of the potential deduced from the particle count (see equation (6.5)), equation (6.8) allows for the multigaussian covariance to be obtained from successive approximations or equivalent graphical method.

F.2 Conditional simulation

Consider a deposit known by a certain number of samples $(v_\alpha)_{\alpha \in A}$. A field D needs to be simulated with the Cox process under the conditions that the number of particles n_α falling in each sample v_α is known. The field D is composed of a union of small disjoint domains v_i of identical size and form as those of the samples.

$$D = \cup_{i \in I}\ v_i$$

with $A \subset I$.

The algorithm consists of three steps:

i. Conditional simulation of the potential of the samples
ii. Conditional simulation of the potential in the field D
iii. Conditional simulation of the particles in the field D.

To avoid complex notations, recall the conventions: Suppose that the family I is totally ordered, something that is allowable. Designate u_I the vector attached to any functions defined on I. As all sub-families of I are also totally ordered, it is equally possible to consider the vector u_A. Finally, if $\alpha \in A$, u_A^α is the vector u_A from which one deducts the component order of α.

Simulation of the potential of the samples

This step consists of the potential of Z_A of the samples under the condition that $N_A = n_A$. Denoting $f(z_A)$ the non-conditional multivariate density of the potential of the informed samples, the conditional distribution of Z_A is proportional to

$$f(z_A) \prod_{\alpha \in A} exp\{-z_\alpha\} \frac{z_\alpha^{n_\alpha}}{n_\alpha!} \tag{6.9}$$

This problem has been studied by Freulon [28, 29] and an iterative solution is proposed, constructed from the Gibbs sampler by Geman and Geman [30].

To initialise Z_A, a method consists of independently simulating each value of the potential taking into account the number of particles in the sample. At the end, an initial vector is generated according to the proportional density of

$$\prod_{\alpha \in A} f(z_\alpha)\ exp\{-z_\alpha\} \frac{z_\alpha^{n_\alpha}}{n_\alpha!} \tag{6.10}$$

where f designates the density of the potential at each sample.

This procedure leads to the simulation of independent gamma distributions if the particle count is negative binomial, or a generalised inverse Gaussian distribution if the particle count is Sichel (described in the appendix of this section).

Let $\alpha_1, \alpha_2, \ldots$ be a sequence of indices taking their value in A in such a way that any value in A occurs infinitely many times in the sequence. This could be a periodic sequence of A, or the realisation of a recurrent, irreducible, discrete time Markov chain of A.

At the iteration l, the algorithm modifies the component of order $\alpha = \alpha_l$ and leaves all the others unchanged. This value z_α is replaced by another value generated according to a distribution proportional to

$$f(z|z_A^\alpha)\ exp\{-z\} \frac{z^{n_\alpha}}{n_\alpha!} \tag{6.11}$$

which requires the density of Z_α given the values taken by the potential of all the other samples. Simulating this conditional density does not present any difficulty in the case where the potential is anamorphosed multigaussian, as everything leads to the simulation of the Gaussian distribution. The acceptance-rejection technique (Von Neumann [66]) gives a procedure to simulate the distribution in (6.11). Having sampled one uniform value u, it follows that a value z, taken according to the conditional density f, is accepted if and only if

$$ln\ u - n_\alpha + n_\alpha\ ln\ n_\alpha \leq z\ +\ n_\alpha\ ln\ z$$

The theory of the Gibbs sampler shows that the law of the vector thus generated converges towards the simulated distribution as the number of simulations becomes very large.

Conditional simulation of the potential in the field

For the simulation of the potential Z_I of the field D given $Z_A = z_A$ of the samples and the number $N_A = n_A$ of the particles that they include**:** The conditional independence of the implantation of the particles makes the condition of particle counts fall away.

Taking into account the anamorphosed, multigaussian model retained for the potential, one finds oneself returning to the classic problem of the conditional simulation of a multigaussian vector Y_I knowing $Y_A = y_A$ with $y_\alpha = \varphi^{-1}(z_\alpha)$ for all $\alpha \in A$.

Conditional simulation of the particles in the field

The conditional independence of the number of particle in disjoint domains allows for a simplifying role. If $i \notin A$, N_i follows the Poisson distribution with mean z_i, independent of values taken by the other samples. If $i = \alpha \in A$, one simply takes $N_i = n_\alpha$.

Conclusion

The use of the Cox process as a simulation model offers the user great flexibility in return for relatively minor restriction in the inference of the model and since it's implementation has proven to be a very successful model to use for the simulation of diamond deposits and associated sample optimisation studies of secondary deposits.

Appendix : Simulation of distributions

Given Z, a positive random variable of density f and given N, a discrete variable which follows a Poisson distribution with mean Z, one has

$$P\{N = n\} = \int_0^{+\infty} e^{-z}\ \frac{z^n}{n!}\ f(z)\ dz$$

and the density of Z knowing $N = n$ is

$$f(z|n) = \frac{e^{-z}\ \frac{z^n}{n!}\ f(z)}{\int_0^{+\infty}\ e^{-z}\ \frac{z^n}{n!}\ f(z)\ dz}$$

Two cases encountered in diamond simulations are examined:

A) If Z follows a Gamma distribution with parameter α and index b

$$f(z) = \frac{b^\alpha}{\Gamma(\alpha)}\ e^{-bz}\ z^{\alpha-1}$$

then N follows a negative binomial distribution with a parameter α.

$$P\{N=n\} = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\ n!}\ (\frac{b}{b+1})^{\alpha}\ (\frac{1}{b+1})^{n}$$

and Z knowing $N=n$ follows a Gamma distribution with parameter $\alpha+n$ and index $b+1$.

B) If Z follows a generalised inverse Gaussian distribution with real parameter γ and positive indices a and b

$$f(z) = \frac{\left(\frac{a}{b}\right)^{\frac{\gamma}{2}}}{2\ K_{\gamma}(2\sqrt{a\ b})}\ z^{\gamma-1}\ exp\{-az\ -\ \frac{b}{z}\}$$

where K_{γ} designates the modified Bessel function of the second kind with indices γ, then N follows a generalised Sichel distribution

$$P\{N=n\} = \frac{1}{n!}\ \frac{\left(\frac{a}{b}\right)^{\frac{\gamma}{2}}}{\left(\frac{a+1}{b}\right)^{\frac{\gamma+n}{2}}}\ \frac{K_{\gamma+n}(2\sqrt{(a+1)b}\ )}{K_{\gamma}(2\sqrt{ab}\ )}$$

and the density of Z knowing $N=n$ is a generalised inverse Gaussian distribution with parameter $\gamma+n$ and indices $a+1$ and b.

* * * * * * * * * * 000 * * * * * * * * * *

Bibliography

[1] Armstrong M. *Basic Linear Geostatistics*. Springer-Verlag, Berlin, 1998.

[2] Assibey-Bonsu W. Summary of present knowledge on the representative sampling of ore in the mining industry. *The Journal of the South African Institute of Mining and Metallurgy*, pages pp.289–293, November 1996.

[3] Barnett WP. *The Mechanics of Kimberlite Emplacement*. PhD thesis, University of KwaZulu Natal, December 2006.

[4] Barnett WP. Fry density targeting. Technical report, De Beers Internal Report, April, 2005.

[5] Boisvert JB and Deutsch CV. Geostatistics with locally varying anisotropy. Technical report, Geocanada, 2009.

[6] Boisvert JB and Deutsch CV. Kriging in the presence of LVA using Dijkstra's algorithm. CCG Membership Report, Tenth Annual Report of the Centre for Computational Geostatistics 2008, University of Alberta, September 2008.

[7] Boisvert JB, Manchuk JG and Deutsch CV. Calculating distance in the Presence of Locally Varying Anisotropy. CCG Membership Report, Ninth Annual Report of the Centre for Computational Geostatistics 2007, University of Alberta, September 2007.

[8] Boisvert JB, Manchuk JG and Deutsch CV. Kriging and Simulation in the Presence of Locally Varying Anisotropy. CCG Membership Report, Ninth Annual Report of the Centre for Computational Geostatistics 2007, University of Alberta, September 2007.

[9] BRC Diamond Core Ltd. website. Diamond recovery process. *http://www.brc-diamondcore.com/s/Technical.asp?ReportID=282505*, Accessed 7 September 2009.

[10] Breiman L, Friedman JH, Olshen RA and Stone CJ. *Classification and Regression Trees*. Chapman & Hall/CRC: Boca Raton, 1998.

[11] Bush D. Effect of Sill and Range on Kriging Weights. Internal Report, De Beers, 1992.

[12] Clark I. Geostatistical estimation and the lognormal distribution. *Geocongress, Pretoria, South Africa*, 1998.

[13] Crough ST, Morgan WJ and Hargraves RB. Kimberlites: Their relation to mantle hotspots. *Earth and Planetary Science Letter, 50*, pages pp.260–274, 1980.

[14] David M. *Development in Geomathematics 2: Geostatistical Ore Reserve Estimation*. Elsevier, Amsterdam, Netherlands, 1977.

[15] Davies R. Glacial model of kimberlite mineral dispersion. Internal Report, Diapros Canada Limited, June 1980.

[16] Davis JC. *Statistics and data analysis in Geology*. John Wiley & Sons Ltd., New York, 2002.

[17] Davy AT. A manual of sample size selection for the diamond industry. Internal Report, De Beers Industrial Diamond Division, 1999.

[18] Dixon PM. Ripley's K function. *Encyclopedia of Environmetrics Vol. 3*, pages 1796–1803, 2002.

[19] Dohm CE. *Improvement of ore evaluation through the identification of homogenous areas using geological, statistical and geostatistical analyses*. Phd Thesis : Wits University, 1995.

[20] Duggan SP. Geostatistical Evaluation of a Marine Deposit. Master of Science Project Report, WITS Univeristy, 1995.

[21] Ebdon D. *Statistics in Geography, Second Edition*. Basil Blackwell, New York, 1988.

[22] Evans M, Hasting N and Peacock B. *Statistical distributions, 3 ed.* John & Wiley, New York, 2000.

[23] Feller W. *An introduction of Probability Theory and its Applications 2*. John & Wiley, New York, 1971.

[24] Ferreira JJ. Tailings granulometry and diamond liberation. Metallurgical Conference, De Beers, 1999.

[25] François-Bongarçon D. Sampling in the mining industry: Theory and Practice. Support document, Seminar: Volume 1: Course notes, 2002.

[26] François-Bongarçon D. The most common error in applying 'Gy's Formula' in the theory of mineral sampling, and the history of the liberation factor. *The Journal of The South African Institute of Mining and Metallurgy*, 2002.

[27] Frempong PK. The development of a robust sampling strategy and protocol in underground gold mines. Technical report, The AusIMM Proceedings, 1999.

[28] Freulon X. *Conditionnement du modéle Gaussian par des inégalités ou des randomisées*. Geostatistics Doctoral Thesis, School of Mines of Paris, 1992.

[29] Freulon X. *Conditional simulation of a Gaussian Random Vector with Non Linear and/or Noisy Observations*. in M Armstrong and PA Dowd (eds.) Geostatistical Simulations, Kluwer, Dortrecht, 1994.

[30] Geman S and Geman D. *Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images*. I.E.E.E. Trans. Pattern Analysis and Machine Intelligence 6, 1984.

[31] Goreaud F and Pélissier R. On explicit formulas of edge effect correction for Ripley's K-function. *Journal of Vegetation Science 10*, pages pp.433–438, 1999.

[32] Grills A, Prins CF, Kilham J and Nordin W. CTD dump sampling design. Internal Presentation, De Beers, August 2003.

[33] Gy P. *Sampling for analytical purposes*. John Wiley & Sons Ltd., Chichester, 1998.

[34] Isaaks E. Local Anisotropy Kriging. Unpublished article, Patent pending, 2005.

[35] Journel AG and Huijbregts Ch J. *Mining Geostatistics*. Academic Press, London, 1978.

[36] Kilham J. Internal De Beers Excel spreadsheet: De Beers Kimberley Mines: CTD-1977-results-xMinSFS.xls. Excel data, De Beers, February 2003.

[37] Kleingeld WJ. *La Geostatistique pour des Variables Discretes*. PhD thesis, Ecole des Mines, Fontainebleau, France, July 1987.

[38] Kleingeld WJ. Dump Sampling Frohe Hoffning - Charlottental. Internal Report, De Beers, December 1977.

[39] Kleingeld WJ and Lantuéjoul C. Sampling of ore bodies with a highly dispersed mineralisation. *Soares A (Ed.) Geostatistics Tróia '92, Kluwer Academic Publishers, Dordrecht*, pages Vol 2. pp 953–964, 1993.

[40] Kleingeld WJ and Prins CF. Tierramar Project. Internal Note, Anglo American, Undated:Approx 1990.

[41] Kleingeld WJ, Lantuéjoul C and Prins CF. *Sampling challenges in highly dispersed types of mineralisation*. The Australasian Institute of Mining and Metallurgy, Proceedings of the Second World Conference on Sampling and Blending, 2005.

[42] Kleingeld WJ, Lantuéjoul C, Prins CF and Thurston ML. The non-conditional simulation of a Cox process with application to a sampling problem. *Proceedings of the Conference on Mining Geostatistics, Berg-en-Dal Conference, South Africa, Geostatistical Association of South Africa*, pages pp. 26–38, September 1994.

[43] Kleingeld WJ, Lantuéjoul C, Thurston ML and Prins CF. The conditional simulation of a Cox process with application to deposits with discrete particles. *Baafi EY and Schofield NA (Eds.) Geostatistics Wollongong '96, Kluwer Academic Publishers, Netherlands*, pages Vol.2 pp 683–694, 1997.

[44] Kleingeld WJ, Lantuéjoul Ch, Prins CF and Thurston ML. The non-conditional simulation of a Cox process with application to a sampling problem. *Proceedings of the Conference on Mining Geostatistics held at Berg-en-Dal Conference Centre, Kruger Park, South Africa*, pages 26–38, 1996.

[45] Lantuéjoul C. On liberation and breakage. Internal Report, De Beers, 1998.

[46] Leuangthong O, Prins CF and Deutsch CV. SGSIM_LVA: Gaussian simulation with locally varying angles. CCG Membership Report, Eighth Annual Report of the Centre for Computational Geostatistics 2006, University of Alberta, September 2006.

[47] Loubser JM. 4 plant dump sampling positions. Internal Report, De Beers, Feb 1998.

[48] Loubser JM. 4 plant dump - minred simulation. Internal Memorandum, De Beers, July 1998.

[49] Matheron G and Armstrong M, editor. *Geostatistical case studies*. D Reidel Publishing Company, Dordrecht, Holland, 1987.

[50] Millad M. A mineral resource estimate for the current tailing mineral resource, Chronofacies 1, Kimberley Mines. Internal Report, De Beers, July 2006.

[51] Olea RA. *Geostatistical Glossary and Multilingual Dictionary*. Oxford University Press, New York, 1991.

[52] Prins CF. Investigating sample size for Douglas Bay prospecting tool. Internal Report, De Beers, April 2003.

[53] Prins CF. Simulating marine deposits using the Cox process. Internal Report, De Beers, February 1995.

[54] Prins CF. De Beers Marine: Research and Progress for 1995 (Working documents). Internal Report, De Beers, March 1996.

[55] Prins CF. Hudson Bay Mining and Smelting: Geostatistical training and working notes. Internal Report, Anglo American, November 1996.

[56] Prins CF and Grills A. Kimberley dump sampling. Internal Presentation, De Beers, July 2003.

[57] Rivoirard J. Two key parameters when choosing the kriging neighbourhood. *Mathematical Geology*, Vol 19:p. 851–856, 1987.

[58] Roex AP. Geochemical correlation between southern African kimberlites and South Atlantic hotspots. *Nature Vol. 324*, pages pp.243–245, 1986.

[59] Royle AG. Alluvial sampling formula and recent advances in alluvial deposit valuation. *Trans. Instn Min. Metall. (Sect. B: Appl. earth sci.)*, pages B179–B182, November 1986.

[60] Sichel HS. Statistical valuation of diamondiferous deposits. *Apcom 1972, Johannesburg*, pages pp.17–24, 1972.

[61] Skinner EMW. Contrasting Group I and Group II kimberlite petrology; Towards a genetic model for kimberlites. *Proceedings of the fourth international kimberlite conference, Volume 1*, pages 528–544, 1986.

[62] Smith CB. Pb, Sr and Nd isotopic evidence for sources of southern Africa Cretaceous kimberlites. *Nature Vol. 304*, pages pp.51–54, 1983.

[63] Spiegel MR. *Theory and Problems of Statistics*. Schaum Outline Series, London, 1992.

[64] Sutherland DG and Dale ML. Method of establishing the minimum sample size for sampling alluvial diamond deposits. *Trans. Instn Min. Metall. (Sect. B: Appl. earth sci.)*, pages B55–B58, May 1984.

[65] Van der Watt P. *Mathematical Statistics: UNISA course notes*. University of South Africa, Course Notes : STW2/1/71-72, 1971.

[66] Von Neumann J. *Various techniques used in connection with random digits*. U.S. Nat. Bur. Stand. Appl. Math. Ser. 12, 1951.

[67] Wikipedia. Gabriel Graph. *Online free encyclopedia*, Accessed 2009.

[68] Wikipedia. Nearest neighbor graph. *Online free encyclopedia*, Accessed 2009.

[69] Zerzan JM. Overlap : A Fortran program for rapidly evaluating the area of overlap between two polygons. *Computers and Geosciences, Vol.15, No.7, Pergamon Press*, pages pp. 1109–1114, December 1989.

* * * * * * * * * * END * * * * * * * * * *

Échantillonnage, simulation et estimation des gisements secondaires de diamant

RÉSUMÉ : Il est difficile d'explorer et d'estimer des gisements secondaires de diamants en raison du manque de fiabilité des données et/ou de leur rareté. Des efforts soutenus sont nécessaires pour maintenir une bonne compréhension de ces types de dépôts lors de leur exploration, leur échantillonnage et leur exploitation. Cette thèse traite des sujets suivants:

(a) L'existence de regroupements entre cheminées kimberlites est établie, et leur extension moyenne déterminée. (b) Des données d'exploration d'indicateurs minéraux sont analysées par arbre de classification. Un modèle est ensuite bâti à partir de sites kimberlitiques connus pour identifier de nouveaux sites. (c) Les milieux maritimes comportent des mécanismes de piégeage complexes, ce qui les rend difficiles à échantillonner. Dans certaines situations, on dispose de peu, voire d'aucune information de qualité, alors qu'une étude d'optimisation de l'échantillonnage est nécessaire. Dans ce cas, une esquisse au crayon est utilisée pour construire des simulations, lesquelles servent à une première optimisation de l'échantillonnage. (d) Dans les dépôts sous-marins profonds, les échantillons doivent dépasser une taille minimale critique pour être représentatifs. L'établissement de cette taille passe par une modélisation selon un processus de Cox, bien adapté à la nature discrète de la minéralisation. L'impact de l'échantillonnage sur la qualité de l'estimation par blocs ou par panneaux peut ainsi être testé par simulation. (e) Ces dernières années, certains terrils sont redevenus économiquement viables. Pour en obtenir une estimation locale par blocs, une procédure de détermination de la taille optimale des échantillons et de leur espacement a été développée et mise en oeuvre sur un terril de kimberlite.

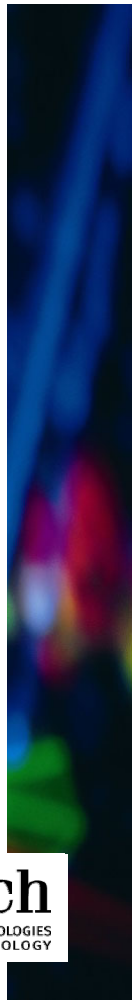
Mots clés : regroupement de kimberlites, arbre de classification, marqueurs minéraux, esquisse au crayon, processus de Cox, échantillonnage des placers, ressources des terrils

Sampling, simulation and estimation of secondary diamond deposits

ABSTRACT : The exploration for diamonds and the evaluation of secondary diamond deposits are challenging processes due to the lack of reliable data and/or the sparsity thereof. Continued effort is required to maintain an understanding of these types of deposits as they are explored, sampled and exploited. This thesis addresses and researches the following:

(a) Demonstrate the clustering of kimberlites and determine the average size of kimberlite targets. (b) A classification tree was applied to exploration indicator minerals and a model using known kimberlite targets was built, then applied to new areas to identify kimberlite targets. (c) Marine environments with complex trapping mechanisms are difficult to sample. In some situations there is very little or even no grade information available but a sample optimisation study is required. A shaded pencil sketch is used and statistically manipulated to create simulations for use as a first approximation to optimise sampling. (d) Sample optimisation studies in deep water marine deposits are well suited to make use of the Cox process to create simulations. The discrete nature of diamonds require a minimum critical size for samples to be representative. Using Cox simulations, the effect of sample size on the accuracy of kriged blocks and panel results is researched. (e) In recent years, some tailings resources have become economical to re-mine. Optimal sample size and spacing of samples and a method to do local block estimates was developed and applied to a kimberlite tailings resource.

Keywords : kimberlite clustering, classification tree, indicator mineral, pencil sketch, Cox process, placer sampling, tailing resources



MINES ParisTech

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY